

TRANSACTIONS  
OF THE  
AMERICAN MATHEMATICAL SOCIETY

EDITED BY

WILLIAM C. GRAUSTEIN

EINAR HILLE

C. C. MAC DUFFEE

WITH THE CO-OPERATION OF

A. A. ALBERT

E. P. LANE

MARSTON MORSE

H. P. ROBERTSON

GABOR SZEGÖ

JESSE DOUGLAS

R. E. LANGER

OYSTEIN ORE

M. H. STONE

G. T. WHYBURN

T. H. HILDEBRANDT

SAUNDERS MACLANE

H. L. RIETZ

J. L. SYNGE

OSCAR ZARISKI

VOLUME 48

JULY TO DECEMBER, 1940

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1940

BOSTON UNIVERSITY  
COLLEGE OF LIBERAL ARTS  
LIBRARY

Chester C. Corbin Library Fund  
Bound Feb. 1941

Composed, Printed and Bound by  
The Edgely Press  
George Banta Publishing Company  
Menasha, Wisconsin

64954



QA  
F001  
v. 48

22

## TABLE OF CONTENTS

### VOLUME 48, JULY TO DECEMBER, 1940

ADAMS, C. R., and MORSE, A. P. Continuous additive functionals on the space $(BV)$ and certain subspaces.....	82
AGNEW, R. P. On kernels of faltung transformations.....	1
BOAS, R. P. Expansions of analytic functions.....	467
BURINGTON, R. S. On circavariant matrices and circa-equivalent networks.....	377
COOLIDGE, J. L. Analytic systems of central conics in space.....	359
DOUGLAS, J. A new special form of the linear element of a surface.....	101
FELLER, W. On the integro-differential equations of purely discontinuous Markoff processes.....	488
GARABEDIAN, H. L., and WALL, H. S. Hausdorff methods of summation and continued fractions.....	185
GLEYZAL, A. Order types and structure of orders.....	451
HALL, D. W., and WHYBURN, G. T. Arc- and tree-preserving transformations.....	63
HALL, M. The position of the radical in an algebra.....	391
JACKSON, D. Orthogonal polynomials with auxiliary conditions.....	72
KASNER, E. Conformality in connection with functions of two complex variables.....	50
MARTIN, W. T. On a minimum problem in the theory of analytic functions of several variables.....	351
MONTGOMERY, D., and ZIPPIN, L. Topological group foundations of rigid space geometry.....	21
MORSE, A. P., and ADAMS, C. R. Continuous additive functionals on the space $(BV)$ and certain subspaces.....	82
NIVEN, I. Integers of quadratic fields as sums of squares.....	405
PHILLIPS, R. S. On linear transformations.....	516
POST, E. L. Polyadic groups.....	208
RITT, J. F. On a type of algebraic differential manifold.....	542
SPENCER, D. C. On finitely mean valent functions. II.....	418
SZÁSZ, O. On strong summability of Fourier series.....	117
TORNHEIM, L. Integral sets of quaternion algebras over a function field.....	436

WALL, H. S. Continued fractions and totally monotone sequences. . . .	165
WALL, H. S., and GARABEDIAN, H. L. Hausdorff methods of summation and continued fractions. . . . .	185
WEYL, H. Theory of reduction for arithmetical equivalence. . . . .	126
WHYBURN, G. T., and HALL, D. W. Arc- and tree-preserving transfor- mations. . . . .	63
ZIPPIN, L., and MONTGOMERY, D. Topological group foundations of rigid space geometry. . . . .	21

# ON KERNELS OF FALTUNG TRANSFORMATIONS

BY

RALPH PALMER AGNEW

1. **Introduction.** A complex-valued function  $J(t)$  defined over  $-\infty < t < \infty$  being given, the function

$$(1.1) \quad y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

is, if it exists, called the *faltung* of the *kernel*  $J(t)$  and the function  $x(t)$ . We use Lebesgue measure and integration, and let  $L$  denote the class of complex-valued functions  $x(t)$  integrable (and hence also absolutely integrable) over the infinite interval  $-\infty < t < \infty$ .

It is well known that if  $J \in L$ , then the faltung  $y(s)$  of each  $x \in L$  exists (that is, is finite) for almost all  $s$ , and  $y \in L$ . This is implied by the computation

$$(1.2) \quad \begin{aligned} \int_{-\infty}^{\infty} y(s)ds &= \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} J(t)x(s+t)dt = \int_{-\infty}^{\infty} J(t)dt \int_{-\infty}^{\infty} x(s+t)ds \\ &= \left[ \int_{-\infty}^{\infty} J(t)dt \right] \left[ \int_{-\infty}^{\infty} x(s)ds \right], \end{aligned}$$

which is justified by the absolute convergence of the integrals involved. If  $J(t)$  is an essentially bounded measurable function, say  $|J(t)| \leq M$  for almost all  $t$ , and  $x \in L$ , then the simple estimate

$$(1.3) \quad \begin{aligned} |y(s)| &\leq \int_{-\infty}^{\infty} |J(t)| |x(s+t)| dt \leq M \int_{-\infty}^{\infty} |x(s+t)| dt \\ &= M \int_{-\infty}^{\infty} |x(t)| dt \end{aligned}$$

shows that  $y(s)$  exists and is bounded over  $-\infty < s < \infty$ . Each of these results is of the type: If  $J$  has property  $P$ , then  $y$  has property  $Q$  for each  $x \in L$ . To supplement such results, it is desirable to know whether the conclusion that  $J$  has property  $P$  can be drawn from the hypothesis that  $y$  has property  $Q$  whenever  $x$  belongs to an appropriate class  $X$  of functions. Doubtless the most pertinent questions are those for which the class  $X$  is  $L$  itself. We are able to obtain affirmative theorems not only when  $X$  is the class  $L$  but also when  $X$  is a suitable class of step functions in  $L$ . Such theorems become stronger and

Presented to the Society, September 7, 1939; received by the editors July 18, 1939.

throw more light on the real character of faltung transformations when the extent of the class  $X$  is reduced. There is some arbitrariness in choice of the classes  $X$ ; we endeavor to make them at the same time as simple and illuminating as possible.

An example may serve to illustrate a role played by step functions in the theory of faltung transformations. If  $J(t) = \exp it^n$ ,  $n > 2$ , then (see §6) simple estimates show that the faltung of each ordinary step function is a bounded continuous function in class  $L$ . But Theorem 3.1 shows that there exist generalized step functions in class  $L$  of which the faltung is not in class  $L$ .

The main results of this paper are Theorems 2.1, 3.1, and 4.1 which are of the following type: If  $y(s)$  has property  $P$  for each  $x(t)$  belonging to a class  $X$  of functions, then  $J(t)$  must have property  $Q$ . With each of these theorems is associated a theorem of familiar type which asserts that, if  $J$  has property  $Q$ , then (i)  $y$  has property  $P$  for each  $x \in L$  and (ii) a certain constant determined by  $J$  is the *bound* of the transformation, that is, the least constant  $M$  such that a constant (norm) determined by  $y$  is less than or equal to  $M \int_{-\infty}^{\infty} |x(t)| dt$  for each  $x \in L$ .

The class  $X$  is in each case a nonlinear subclass of  $L$  consisting of certain generalized non-negative step functions. Neither the class  $X$ , nor the larger manifold  $\mathfrak{M}(X)$  consisting of all finite linear combinations of elements of  $X$ , forms a closed set in the space  $L$  in which the distance between two elements  $x_1(t)$  and  $x_2(t)$  of  $L$  is given by the familiar metric

$$(1.4) \quad \int_{-\infty}^{\infty} |x_2(t) - x_1(t)| dt;$$

in other words the space obtained by using the elements of  $\mathfrak{M}(X)$  and the metric of  $L$  is not complete. It is shown in §6 that each of Theorems 2.1, 3.1, and 4.1 will fail if  $X$  is replaced by certain smaller classes of step functions.

Let  $S$  denote the special class of all real non-negative functions  $x = x(t)$  such that (i)  $x \in L$  and (ii) there exist non-negative constants  $\dots, c_{-1}, c_0, c_1, c_2, \dots$  and  $\dots < a_{-1} < a_0 < a_1 < a_2 < \dots$  (depending on the particular function  $x$ ) such that  $\lim_{n \rightarrow -\infty} a_n = -\infty$ ,  $\lim_{n \rightarrow \infty} a_n = \infty$ , and for each  $n = \dots, -1, 0, 1, 2, \dots$ ,

$$(1.5) \quad x(t) = c_n, \quad a_n \leq t < a_{n+1}.$$

Each  $x \in S$  may be described as a real non-negative function in  $L$  which is a generalized step function<sup>(1)</sup> having a finite number of steps in each finite interval.

Let  $S_U$  denote the subclass of  $S$  consisting of those functions in  $S$  for which  $a_{n+1} - a_n = 1$ ,  $n = 0, \pm 1, \pm 2, \dots$ ; each  $x \in S_U$  is a *unit step function*, each step

<sup>(1)</sup> We reserve the term *ordinary step function* for step functions which vanish outside some finite interval.

being one unit long. Each  $x \in S$  is bounded over each finite interval, and each  $x \in S_U$  is bounded over  $-\infty < t < \infty$ . (It is a trivial remark that the last assertion would be false if in (1.5)  $a_n \leq t < a_{n+1}$  were replaced by  $a_n < t < a_{n+1}$ .)

2. **Conditions for existence of  $y(s)$ .** This section is devoted to discussion and proof of the following two theorems.

**THEOREM 2.1.** *If  $J(t)$  is such that, for each  $x \in S_U$ ,*

$$(2.11) \quad y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

*exists for at least one  $s$  in the interval  $-\infty < s < \infty$ , then  $J(t)$  is measurable<sup>(\*)</sup> over  $-\infty < t < \infty$  and for each constant  $0 < A < \infty$  there is a constant  $M_A$  such that*

$$(2.12) \quad \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+A} |J(t)| dt = M_A < \infty.$$

**THEOREM 2.2.** *If  $J(t)$  is measurable and (2.12) holds, then for each  $x \in L$ ,  $y(s)$  defined by (2.11) exists for almost all  $s$  and is measurable, and for each  $A > 0$*

$$(2.21) \quad \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+A} |y(s)| ds \leq M_A \int_{-\infty}^{\infty} |x(t)| dt$$

*where  $M_A$  is the constant of (2.12). Moreover the constant  $M_A$  in (2.21) is the best possible one in the sense that if a measurable function  $J(t)$  satisfying (2.12) and  $A > 0$  are fixed, then, for each  $C < M_A$ ,*

$$(2.22) \quad \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+A} |y(s)| ds > C \int_{-\infty}^{\infty} |x(t)| dt$$

*will be true for some  $x \in L$ .*

If  $A_1$  and  $A_2$  are finite positive numbers, then each interval  $u \leq t \leq u + A_1$  can be covered by a finite set of intervals of the form  $u_k \leq t \leq u_k + A_2$ ; hence it is apparent that if the left member of (2.12) is finite for some one  $A > 0$ , then it is finite for each  $A > 0$ . Therefore the condition (2.12) is equivalent to

$$(2.23) \quad \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+1} |J(t)| dt < \infty,$$

and this condition is easily seen to be equivalent to

$$(2.24) \quad \text{l.u.b.}_{n=0, \pm 1, \pm 2, \dots} \int_n^{n+1} |J(t)| dt < \infty.$$

(\*) Perhaps little would be lost if we were to assume measurability of  $J(t)$ ; but the proof of measurability of  $J(t)$  is so simple (see the few lines following the statement of Lemma 2.3) that we elect to prove it rather than to assume it.

In §7 we discuss further the class of functions satisfying the inequality (2.12).

It is a corollary of Theorems 2.1 and 2.2 that if  $J$  is such that  $y(s)$  exists for at least one  $s$  whenever  $x \in L$ , then  $y(s)$  must exist for almost all  $s$  whenever  $x \in L$ . This does not imply that if  $J(t)$  and  $x(t)$  are a pair of functions with  $x \in L$  for which  $y(s)$  exists for at least one  $s$ , then  $y(s)$  must exist for almost all  $s$ ; indeed if  $J(t)$  is 0 or  $t^2$  according as  $[t]$ , the greatest integer less than or equal to  $t$ , is even or odd, and  $x(t)$  is  $1/(1+t^2)$  or 0 according as  $[t]$  is even or odd, then  $x \in L$  and  $y(s)=0$  when  $s$  is an integer but  $y(s)=\infty$  when  $s$  is not an integer.

Our first step in the proof of Theorem 2.1 is to prove

LEMMA 2.3. *If  $J(t)$  is such that, for each  $x \in S_U$ ,  $y(s)$  exists for at least one  $s$ , then  $J(t)$  is integrable over each finite interval  $a \leq t \leq b$  and, for each  $x \in S_U$ ,  $y(s)$  exists for at least one  $s$  in each closed interval of unit length in the interval  $-\infty < s < \infty$ .*

To prove Lemma 2.3, let  $J(t)$  satisfy its hypothesis and let  $x_0(t)$  be a positive function  $x \in S_U$ , say  $x_0(t) = 1/(1+[t]^2)$ . Let  $s_0$  be fixed such that

$$(2.31) \quad y_0(s_0) = \int_{-\infty}^{\infty} J(t)x_0(s_0+t)dt$$

exists. Then if  $-\infty < a < b < \infty$ ,

$$\int_a^b J(t)x_0(s_0+t)dt$$

exists. But the function  $1/x_0(s_0+t)$  is measurable and bounded over  $a \leq t \leq b$ ; therefore  $J(t) = [J(t)x_0(s_0+t)]/x_0(s_0+t)$  is integrable as well as measurable over  $a < t < b$ .

Now let an arbitrary function  $x \in S_U$  be fixed. The function  $X(t)$  defined by the series

$$(2.32) \quad X(t) = \sum_{n=-\infty}^{\infty} 2^{-|n|}x(t+n)$$

exists for almost all  $t$ , and  $X \in S_U$ ; that  $X \in L$  is shown by the computation

$$\begin{aligned} \int_{-\infty}^{\infty} X(t)dt &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} 2^{-|n|}x(t+n)dt \\ &= \sum_{n=-\infty}^{\infty} 2^{-|n|} \int_{-\infty}^{\infty} x(t+n)dt = 3 \int_{-\infty}^{\infty} x(t)dt \end{aligned}$$

which is justified by the fact that  $x(t) \geq 0$  and  $x \in L$ ; and  $X(t)$  and  $x(t)$  are constant over the same unit intervals. Let  $s_0$  be fixed such that



$$(2.33) \quad Y(s_0) = \int_{-\infty}^{\infty} J(t)X(s_0 + t)dt$$

exists. Then

$$(2.34) \quad \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} 2^{-|n|} |J(t)| x(s_0 + n + t)dt$$

exists, and since each term in the sum is measurable and non-negative, this implies that

$$(2.35) \quad \int_{-\infty}^{\infty} |J(t)| x(s_0 + n + t)dt$$

exists for each  $n$ . Thus

$$(2.36) \quad y(s) = \int_{-\infty}^{\infty} J(t)x(s + t)dt$$

exists when  $s = s_0, s_0 \pm 1, s_0 \pm 2, \dots$ . Since each closed unit interval contains at least one of these points, Lemma 2.3 is proved.

To complete the proof of Theorem 2.1, let  $J(t)$  satisfy the hypothesis of Theorem 2.1 and hence the conclusion of Lemma 2.3. To establish (2.12), we assume that (2.12) fails and obtain a contradiction. Failure of (2.12) implies that the left member of (2.24) is  $+\infty$ ; hence there is a sequence  $n_1, n_2, n_3, \dots$  of integers such that  $|n_p - n_q| > 3, p \neq q$ , and

$$(2.37) \quad \lim_{\alpha \rightarrow \infty} I(n_\alpha) = \infty$$

where

$$I(n) = \int_n^{n+1} |J(t)| dt.$$

It follows from (2.37) that we can choose a decreasing sequence  $\theta_1 > \theta_2 > \dots$  of positive numbers such that

$$(2.38) \quad \sum_{\alpha=1}^{\infty} I(n_\alpha)\theta_\alpha = \infty, \quad \sum_{\alpha=1}^{\infty} \theta_\alpha < \infty.$$

Let

$$(2.39) \quad \begin{aligned} x(t) &= \theta_\alpha, & n_\alpha - 1 \leq t < n_\alpha + 2, \alpha = 1, 2, \dots, \\ &= 0, & \text{otherwise.} \end{aligned}$$

Then  $x(t)$  is real, non-negative, and constant over each of the abutting unit intervals  $n \leq t < n+1$ ; and the second of the relations (2.38) implies that  $x \in L$ .

Hence  $x \in S_U$ . Since the integrands are all measurable and non-negative, we find when  $|s| \leq 1$

$$\begin{aligned} \int_{-\infty}^{\infty} |J(t)| |x(s+t)| dt &= \int_{-\infty}^{\infty} |J(t-s)| |x(t)| dt \\ &= \sum_{\alpha=1}^{\infty} \int_{3n_{\alpha}-1}^{n_{\alpha}+2} |J(t-s)| |x(t)| dt \\ &= \sum_{\alpha=1}^{\infty} \theta_{\alpha} \int_{3n_{\alpha}-1}^{n_{\alpha}+2} |J(t-s)| dt \\ &\geq \sum_{\alpha=1}^{\infty} \theta_{\alpha} \int_{n_{\alpha}}^{n_{\alpha}+1} |J(t)| dt = \sum_{\alpha=1}^{\infty} \theta_{\alpha} I(n_{\alpha}) = \infty; \end{aligned}$$

hence

$$y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

fails to exist when  $|s| \leq 1$  and we have a contradiction of the fact that  $y(s)$  must exist for at least one  $s$  in each unit interval. This completes the proof of Theorem 2.1.

Proof of the first part of Theorem 2.2 is very simple. Assuming that  $J(t)$  is measurable and (2.12) holds, and that  $x \in L$  and  $A > 0$  are fixed, we find for each real  $u$

$$\begin{aligned} \int_u^{u+A} |y(s)| ds &\leq \int_u^{u+A} ds \int_{-\infty}^{\infty} |J(t)x(s+t)| dt \\ &= \int_u^{u+A} ds \int_{-\infty}^{\infty} |J(t-s)| |x(t)| dt \\ &= \int_{-\infty}^{\infty} |x(t)| dt \int_u^{u+A} |J(t-s)| ds \\ &= \int_{-\infty}^{\infty} |x(t)| dt \int_{t-u-A}^{t-u} |J(\alpha)| d\alpha \leq M_A \int_{-\infty}^{\infty} |x(t)| dt; \end{aligned}$$

the steps are easily justified by fundamental theorems which imply also that  $y(s)$  exists almost everywhere and is measurable over  $u \leq s \leq u+A$ . It follows immediately that  $y(s)$  exists almost everywhere and is measurable over  $-\infty < s < \infty$ ; and that (2.21) holds.

In our proof of the last part of Theorem 2.2 we shall use the following lemma in which we choose notation to fit the application.

**LEMMA 2.4.** *If  $u$  is real,  $A > 0$ ,  $h > 0$ , and  $J(t)$  is integrable over the interval  $u \leq t \leq u+A+h$ , then*



$$(2.41) \quad \lim_{\delta \rightarrow 0+} \int_u^{u+A} ds \frac{1}{\delta} \int_s^{s+\delta} |J(t) - J(s)| dt = 0.$$

In case  $J(t)$  is continuous over  $u \leq t \leq u+A+h$  we can, for each  $\epsilon > 0$ , choose  $\delta_0 > 0$  such that  $|J(t) - J(s)| < \epsilon/A$  when  $s$  and  $t$  lie between  $u$  and  $u+A+\delta_0$  and  $|t-s| < \delta_0$ ; letting  $I(\delta)$  denote the iterated integral in (2.41), we find that  $0 \leq I(\delta) < \epsilon$  when  $0 < \delta < \delta_0$  and (2.41) follows. In case  $J(t)$  is not continuous, we can show that  $\limsup_{\delta \rightarrow 0} I(\delta) < \epsilon$  by use of the following inequality:

$$|J(t) - J(s)| \leq |J(t) - J_s(t)| + |J_s(t) - J_s(s)| + |J_s(s) - J(s)|$$

in which  $J_s(t)$  is a function continuous over  $u \leq t \leq u+A+h$  for which

$$\int_u^{u+A+h} |J(t) - J_s(t)| dt < \epsilon/3.$$

Let a measurable function  $J(t)$  for which (2.12) holds and a constant  $A > 0$  be fixed. For each  $\delta > 0$ , let  $x_\delta(t)$  be defined by

$$(2.42) \quad \begin{aligned} x_\delta(t) &= \delta^{-1}, & 0 \leq t < \delta, \\ &= 0, & \text{otherwise.} \end{aligned}$$

Then

$$(2.43) \quad \int_{-\infty}^{\infty} |x_\delta(t)| dt = 1, \quad \delta > 0.$$

The faltung  $y_\delta(s)$  of  $J$  and  $x_\delta$  is

$$y_\delta(s) = \int_{-\infty}^{\infty} J(t) x_\delta(s+t) dt = \frac{1}{\delta} \int_s^{s+\delta} J(t) dt;$$

hence

$$y_\delta(-s) - J(s) = \frac{1}{\delta} \int_s^{s+\delta} [J(t) - J(s)] dt$$

so that

$$|y_\delta(-s) - J(s)| \leq \frac{1}{\delta} \int_s^{s+\delta} |J(t) - J(s)| dt$$

and for each  $u$

$$\int_u^{u+A} |y_\delta(-s) - J(s)| ds \leq \int_u^{u+A} ds \frac{1}{\delta} \int_s^{s+\delta} |J(t) - J(s)| dt.$$

Using Lemma 2.3, we obtain

$$\lim_{\delta \rightarrow 0} \int_u^{u+A} |y_\delta(-s) - J(s)| ds = 0,$$

and this implies that

$$(2.44) \quad \lim_{\delta \rightarrow 0} \int_u^{u+A} |y_\delta(-s)| ds = \int_u^{u+A} |J(s)| ds.$$

If now  $C < M_A$ , then we can choose a fixed  $u$  such that

$$(2.45) \quad \int_u^{u+A} |J(s)| ds > C,$$

and then because of (2.44) and (2.45) we can choose a fixed  $\delta > 0$  such that

$$(2.46) \quad \int_u^{u+A} |y_\delta(-s)| ds > C.$$

Using (2.42), we can write (2.46) in the form

$$(2.47) \quad \int_{-u-A}^{-u} |y_\delta(s)| ds > C \int_{-\infty}^{\infty} |x_\delta(t)| dt;$$

this implies (2.22) and Theorem 2.2 is proved.

The hypotheses of Theorems 2.1 and 2.2 do not imply that, if  $x \in L$ , then  $y(s)$  must exist for all real  $s$ . This follows from

**THEOREM 2.5.** *In order that  $J(t)$  may be such that*

$$y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

*exists for all real  $s$  whenever  $x \in L$ , it is necessary and sufficient that  $J(t)$  be measurable and essentially bounded.*

A function  $J(t)$  is called essentially bounded if there is a constant  $M$  such that  $|J(t)| < M$  for almost all  $t$ . Sufficiency is a consequence of the well known fact that if  $J(t)$  is measurable and essentially bounded and  $\xi(t) \in L$ , then  $J(t)\xi(t) \in L$ ; and necessity is a consequence of the well known fact that if  $J(t)\xi(t) \in L$  for each  $\xi \in L$ , then  $J(t)$  is measurable and essentially bounded. If  $J(t)$  is essentially bounded, then (2.12) holds and  $M_A \leq A\beta$  where  $\beta$  is the least constant such that  $|J(t)| \leq \beta$  for almost all  $t$ ; but (2.12) does not imply that  $J(t)$  is essentially bounded.

**3. Conditions for  $y \in L$ .** It is possible to prove, by means of an extension of a theorem of Banach<sup>(3)</sup> and some ideas which we use in the course of proof

<sup>(3)</sup> Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 87, Theorem 9. The extension required is from the finite interval  $0 \leq t \leq 1$  to the infinite interval  $-\infty < t < \infty$ , and from real-valued functions to complex-valued functions.

of Theorem 3.1, that if  $J(t)$  is such that, for each  $x \in L$ ,  $y(s)$  exists for almost all  $s$  and  $y \in L$ , then  $J \in L$ . Theorem 3.1 below, of which we give a direct proof, includes this result.

THEOREM 3.1. *If  $J(t)$  is such that, for each  $x \in S$ ,*

$$(3.11) \quad y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

*exists for almost all  $s$  and  $y \in L$ , then  $J \in L$ .*

THEOREM 3.2. *If  $J \in L$ , then for each  $x \in L$ ,  $y(s)$  as defined by (3.11) exists for almost all  $s$ ,  $y \in L$ , and*

$$(3.21) \quad \int_{-\infty}^{\infty} |y(s)| ds \leq M_{\infty} \int_{-\infty}^{\infty} |x(t)| dt$$

*where*

$$(3.22) \quad M_{\infty} = \int_{-\infty}^{\infty} |J(t)| dt;$$

*moreover  $M_{\infty}$  is the best possible constant in (3.21) in the sense that if  $C < M_{\infty}$  then*

$$(3.23) \quad \int_{-\infty}^{\infty} |y(s)| ds > C \int_{-\infty}^{\infty} |x(t)| dt$$

*will hold for some  $x \in L$ .*

Our first step in the proof of Theorem 3.1 is to prove

LEMMA 3.3. *If  $J(t)$  is such that  $y \in L$  whenever  $x \in S$ , then there is a constant  $M < \infty$  such that*

$$(3.31) \quad \int_{-\infty}^{\infty} |y(s)| ds \leq M \int_{-\infty}^{\infty} |x(t)| dt, \quad x \in S_1,$$

*where  $S_1$  is the subclass of  $S$  consisting of those functions  $x(t)$  in  $S$  which vanish outside the interval  $0 \leq t < 1$ .*

If  $J(t)$  satisfies the hypothesis of Lemma 3.3, and no  $M < \infty$  exists for which (3.31) holds, then for each  $n = 1, 2, \dots$  there is  $x_n \in S_1$  such that

$$(3.32) \quad \int_{-\infty}^{\infty} |y_n(s)| ds > 2^n \int_{-\infty}^{\infty} |x_n(t)| dt,$$

$y_n$  being the transform of  $x_n$ . Since the faltung transformation is homogeneous, we can assume that the functions  $x_n(t)$  and  $y_n(t)$  are divided by the left member of (3.32) so that

$$(3.33) \quad \int_{-\infty}^{\infty} |x_n(t)| dt < 2^{-n}, \quad \int_{-\infty}^{\infty} |y_n(s)| ds = 1.$$

Let  $\lambda_1 = 0$  and choose constants  $a_2 > a_1 + 1$  such that the inequality

$$(3.34) \quad \int_{a_n}^{a_{n+1}} |y_n(s - \lambda_n)| ds > 1 - 3^{-n}$$

holds when  $n = 1$ . Since

$$(3.35) \quad \lim_{\lambda_{n+1} \rightarrow \infty} \int_{a_{n+1}}^{\infty} |y_{n+1}(s - \lambda_{n+1})| ds = 1,$$

we can choose  $\lambda_2 > \lambda_1 + 1$  and then choose  $a_3 > a_2 + 1$  in such a way that (3.34) holds when  $n = 2$ . Continuation by induction furnishes sequences  $a_1 < a_2 < a_3 < \dots$  and  $\lambda_1 < \lambda_2 < \lambda_3 < \dots$  such that  $a_{n+1} > a_n + 1$ ,  $\lambda_{n+1} > \lambda_n + 1$ , and (3.34) holds for each  $n = 1, 2, \dots$ . Since  $x_n(t) \in S_1$ , it follows that  $x_n(t - \lambda_n)$  vanishes outside the interval  $\lambda_n \leq t \leq \lambda_n + 1$ . Let

$$(3.36) \quad X(t) = \sum_{n=1}^{\infty} x_n(t - \lambda_n).$$

The series converges for each  $t$  since, for each  $t$ ,  $x_n(t - \lambda_n) \neq 0$  for at most one  $n$ . Properties of the sequences  $x_n$  and  $\lambda_n$  imply that  $X \in S$ . Hence, by hypothesis,

$$(3.37) \quad Y(s) = \int_{-\infty}^{\infty} J(t)X(s+t)dt$$

exists for almost all  $s$ , and  $Y \in L$ . Since  $X(t)$  vanishes at all points  $t$  not in one of the mutually exclusive intervals  $(\lambda_n, \lambda_n + 1)$ , it follows from (3.37) that

$$\begin{aligned} Y(s) &= \sum_{n=1}^{\infty} \int_{\lambda_n}^{\lambda_n+1} J(t-s)X(t)dt \\ (3.38) \quad &= \sum_{n=1}^{\infty} \int_{\lambda_n}^{\lambda_n+1} J(t-s)x_n(t-\lambda_n)dt = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} J(t-s)x_n(t-\lambda_n)dt \\ &= \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} J(t)x_n(s-\lambda_n+t)dt = \sum_{n=1}^{\infty} y_n(s-\lambda_n). \end{aligned}$$

It follows from (3.33) and (3.34) that for each  $n = 1, 2, \dots$

$$(3.39) \quad \int_{a_n}^{a_{n+1}} |y_k(s - \lambda_k)| ds > 1 - 3^{-n}, \quad k = n, \\ < 3^{-k}, \quad k \neq n.$$

Hence the inequality

$$(3.41) \quad |Y(s)| = \left| \sum_{k=1}^{\infty} y_k(s - \lambda_k) \right| \geq |y_n(s - \lambda_n)| - \sum_{k \neq n} |y_k(s - \lambda_k)|$$

implies that

$$(3.42) \quad \int_{a_n}^{a_{n+1}} |Y(s)| ds \geq 1 - 3^{-n} - \sum_{k \neq n} 3^{-k} = \frac{1}{2}, \quad n = 1, 2, \dots$$

This is inconsistent with the previous conclusion that  $Y \in L$ ; hence  $M < \infty$  exists for which (3.31) holds and Lemma 3.3 is proved.

To prove Theorem 3.1, let  $J(t)$  satisfy its hypothesis, and let  $x_\delta(t)$  be the function in (2.42) which is  $\delta^{-1}$  over  $0 \leq t < \delta$  and is 0 otherwise. If  $0 < \delta < 1$ , then  $x_\delta \in S_1$ ; hence Lemma 3.3 implies existence of a constant  $D < \infty$  such that

$$(3.43) \quad \int_{-\infty}^{\infty} |y_\delta(s)| ds \leq D \int_{-\infty}^{\infty} |x_\delta(s)| ds = D, \quad 0 < \delta < 1.$$

Since, by Theorem 2.1,  $J(t)$  is integrable over each finite interval, (2.44) must hold; replacing  $A$  by  $2A$  and setting  $u = -A$  in (2.44) gives

$$(3.44) \quad \int_{-A}^A |J(t)| dt = \lim_{\delta \rightarrow 0} \int_{-A}^A |y_\delta(-s)| ds, \quad A > 0.$$

From (3.43) and (3.44) we obtain

$$(3.45) \quad \int_{-A}^A |J(t)| dt \leq D, \quad A > 0,$$

and this implies that

$$(3.46) \quad \int_{-\infty}^{\infty} |J(t)| dt \leq D.$$

Thus  $J \in L$  and Theorem 3.1 is proved.

The first part of Theorem 3.2 is well known, and we give its proof merely for completeness. If  $J \in L$  and  $x \in L$ , the computation

$$(3.47) \quad \begin{aligned} \int_{-\infty}^{\infty} |y(s)| ds &\leq \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} |J(t)| |x(s+t)| dt \\ &= \int_{-\infty}^{\infty} |J(t)| dt \int_{-\infty}^{\infty} |x(s+t)| dt \\ &= \left[ \int_{-\infty}^{\infty} |J(t)| dt \right] \left[ \int_{-\infty}^{\infty} |x(s)| ds \right] \end{aligned}$$

is easily justified and (3.21) follows. If it be assumed that

$$(3.48) \quad \int_{-\infty}^{\infty} |y(s)| ds \leq C \int_{-\infty}^{\infty} |x(t)| dt, \quad x \in L,$$

where

$$(3.49) \quad C < \int_{-\infty}^{\infty} |J(t)| dt,$$

then we can set  $D = C$  in (3.43) to obtain  $D = C$  in (3.46) and have a contradiction of (3.49). Therefore if  $C < M_{\infty}$ , then  $x \in L$  exists for which (3.23) holds and Theorem 3.2 is proved.

**4. Conditions for  $y \in B$ .** The measurable upper bound of a real measurable function  $\xi(t)$  defined over  $-\infty < t < \infty$  is the least number  $\beta$  such that  $\xi(t) \leq \beta$  for almost all  $t$ . We write  $\beta = \text{m.u.b. } \xi(t)$ ; and let  $B$  denote the class of all complex-valued measurable functions  $x(t)$  for which  $\text{m.u.b. } |x(t)| < \infty$ .

**THEOREM 4.1.** *If  $J(t)$  is such that, for each  $x \in S$ ,*

$$(4.11) \quad y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

*exists for almost all  $s$  and  $y \in B$ , then  $J \in B$ .*

**THEOREM 4.2.** *If  $J \in B$ , then for each  $x \in L$ ,  $y(s)$  as defined by (4.11) exists for all  $s$  and*

$$(4.21) \quad \text{l.u.b.}_{-\infty < s < \infty} |y(s)| \leq \beta \int_{-\infty}^{\infty} |x(t)| dt$$

where

$$(4.22) \quad \beta = \text{m.u.b.}_{-\infty < t < \infty} |J(t)|;$$

moreover  $\beta$  is the best possible constant in (4.21) in the sense that, if  $C < \beta$ , then

$$(4.23) \quad \text{l.u.b.}_{-\infty < s < \infty} |y(s)| > C \int_{-\infty}^{\infty} |x(t)| dt$$

will hold for some  $x \in L$ .

Our first step in the proof of Theorem 4.1 is to prove the following lemma in which  $S$  and  $S_1$  denote the classes of step functions previously defined in §1 and Lemma 3.3.

**LEMMA 4.3.** *If  $J(t)$  is such that  $y \in B$  whenever  $x \in S$ , then there is a constant  $M$  such that*

$$(4.31) \quad \text{m.u.b.}_{-\infty < s < \infty} |y(s)| \leq M \int_{-\infty}^{\infty} |x(t)| dt, \quad x \in S_1.$$

To prove Lemma 4.3, let  $J(t)$  satisfy its hypothesis and assume that (4.31) fails. Failure of (4.31) implies existence for each  $n = 1, 2, 3, \dots$  of a function  $x_n \in S_1$  having a transform  $y_n$  such that

$$\text{m.u.b.}_{-\infty < s < \infty} |y_n(s)| > 4^n \int_{-\infty}^{\infty} |x_n(t)| dt.$$

We can suppose that each  $x_n$ , and hence also  $y_n$ , has been multiplied by the appropriate constant to give

$$(4.32) \quad \text{m.u.b.}_{-\infty < s < \infty} |y_n(s)| = 2^n; \quad \int_{-\infty}^{\infty} |x_n(t)| dt \leq 2^{-n}, \quad n = 1, 2, \dots$$

If  $\theta_1, \theta_2, \dots$  is a sequence of which each element is either 0 or 1, then

$$(4.33) \quad X(t) \equiv \sum_{n=1}^{\infty} \theta_n x_n(t-n) \in S.$$

Hence under our hypothesis

$$(4.34) \quad Y(s) \equiv \int_{-\infty}^{\infty} J(t)X(s+t)dt \in B.$$

Starting with (4.34), we obtain

$$(4.35) \quad Y(s) = \sum_{n=1}^{\infty} \theta_n y_n(s-n) \in B.$$

That the conclusions just obtained are contradictory, and hence that Lemma 4.3 is true, is a consequence of the following lemma in which we write  $w_n(s)$  for  $y_n(s-n)$ .

LEMMA 4.4. *If  $w_n(s)$  is a sequence of measurable functions, defined over  $-\infty < s < \infty$ , such that*

$$(4.41) \quad \text{m.u.b.}_{-\infty < s < \infty} \left| \sum_{n=1}^{\infty} \theta_n w_n(s) \right| < \infty$$

*for each sequence  $\theta_n$  of which each element is 0 or 1, then there is a constant  $Q < \infty$  such that*

$$(4.42) \quad \text{m.u.b.}_{-\infty < s < \infty} |w_n(s)| \leq Q, \quad n = 1, 2, 3, \dots$$

If  $p$  is a positive integer and we set  $\theta_n = 0$  or 1 according as  $n \neq p$  or  $n = p$ , we see that a constant  $Q_p < \infty$  exists such that

$$(4.43) \quad \text{m.u.b.}_{-\infty < s < \infty} |w_p(s)| = Q_p.$$

To prove (4.42) amounts to proving that the sequence  $Q_p$  is bounded. As-



sume to the contrary that

$$(4.44) \quad \limsup_{p \rightarrow \infty} Q_p = \infty.$$

Setting  $\theta_n = 1$  for each  $n$  in (4.41) shows that the series  $\sum_{n=1}^{\infty} w_n(s)$  converges for almost all  $s$ ; hence  $\lim_{n \rightarrow \infty} w_n(s) = 0$  for almost all  $s$ . Therefore by a theorem of Egoroff<sup>(4)</sup>  $w_n(s)$  converges to 0 essentially uniformly over each set of finite measure  $|E|$ ; that is, corresponding to each  $\delta > 0$ , there is a subset  $F$  of  $E$  such that  $|E - F| < \delta$  and  $w_n(s)$  converges to 0 uniformly over  $F$ . Using (4.44), choose an index  $n_1$  such that  $R_1 \equiv \text{m.u.b. } |w_n(s)| > 2$  when  $n = n_1$ . Let  $E_1$  be a bounded set of positive measure such that  $|w_n(s)| > 1$  when  $n = n_1$ ,  $s \in E_1$ . Let  $F_1$  be a subset of  $E_1$  such that  $|E_1 - F_1| < |E_1|/2^2$  and  $w_n(s)$  converges to 0 uniformly over  $F_1$ . Choose  $n_2 > n_1$  such that  $|w_n(s)| < 2^{-2}$  when  $n = n_2$ ,  $s \in F_1$  and also  $R_2 \equiv \text{m.u.b. } |w_n(s)| > 3 + R_1$  when  $n = n_2$ . Let  $E_2$  be a bounded set of positive measure such that  $|w_n(s)| > 2 + R_1$  when  $n = n_2$ ,  $s \in E_2$ . Let  $F_2$  be a subset of  $E_1 + E_2$  such that  $|E_1 - F_2| < |E_1|/2^3$ ,  $|E_2 - F_2| < |E_2|/2^3$ , and such that  $w_n(s)$  converges to 0 uniformly over  $F_2$ . Choose  $n_3 > n_2$  such that  $|w_n(s)| < 2^{-3}$  when  $n = n_3$ ,  $s \in F_2$ , and also  $R_3 \equiv \text{m.u.b. } |w_n(s)| > 4 + R_1 + R_2$  when  $n = n_3$ . We continue by induction to obtain sequences of numbers and sets such that for each  $p = 1, 2, 3, \dots$

$$(4.45) \quad |E_p| > 0; \quad \text{m.u.b. } |w_{n_p}(s)| = R_p; \quad -\infty < s < \infty$$

$$(4.46) \quad |w_{n_p}(s)| > p + \sum_{k=1}^{p-1} R_k, \quad s \in E_p;$$

$$(4.47) \quad F_p \subset \sum_{k=1}^p E_k; \quad |w_{n_p}(s)| < 2^{-p}, \quad s \in F_{p-1};$$

$$(4.48) \quad |E_k - F_p| \leq |E_k|/2^{p+1}, \quad k \leq p.$$

Setting, for each positive integer  $k$ ,  $G_k = E_k F_k F_{k+1} F_{k+2} \dots$ , we find

$$E_k - G_k = E_k - \prod_{p=k}^{\infty} F_p = \sum_{p=k}^{\infty} (E_k - F_p)$$

so that by (4.48)

$$|E_k - G_k| \leq \sum_{p=k}^{\infty} |E_k - F_p| \leq \sum_{p=k}^{\infty} |E_k|/2^{p+1} \leq |E_k|/2;$$

and therefore  $|G_k| \geq |E_k|/2 > 0$ . Let

$$W(s) = \sum_{p=1}^{\infty} w_{n_p}(s).$$

<sup>(4)</sup> See, for example, E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. 2, p. 144. The extension of the theorem to complex-valued functions is easily made.



For almost all  $s$  in  $G_k$  we find on using (4.45), (4.46), and (4.47) that

$$\begin{aligned} |W(s)| &\geq - \sum_{p=1}^{k-1} |w_{n_p}(s)| + |w_{n_k}(s)| - \sum_{p=k+1}^{\infty} |w_{n_p}(s)| \\ &\geq - \sum_{p=1}^{k-1} R_p + k + \sum_{p=1}^{k-1} |R_p| - \sum_{p=k+1}^{\infty} 2^{-p} \geq k-1. \end{aligned}$$

Hence for each integer  $k$  there is a set of measure  $|G_k| > 0$  such that  $|W(s)| \geq k-1$  for all  $s$  in the set. Therefore

$$(4.49) \quad \text{m.u.b.} \left| \sum_{p=1}^{\infty} w_{n_p}(s) \right| = \infty.$$

But (4.49) contradicts the hypothesis that (4.41) must hold in case  $\theta_n$  is 1 when  $n = n_1, n_2, n_3, \dots$  and 0 otherwise. This completes the proof of Lemma 4.4 and hence also that of Lemma 4.3.

To prove Theorem 4.1, let  $J(t)$  satisfy its hypothesis. By Lemma 4.3 there is a constant  $D < \infty$  such that

$$(4.51) \quad \text{m.u.b.} \left| \int_{-\infty}^{\infty} J(t)x(s+t)dt \right| \leq D \int_{-\infty}^{\infty} |x(t)| dt, \quad x \in S_1.$$

Let, where  $0 < \delta < 1$ ,  $x_{\delta}(t) = \delta^{-1}$  when  $0 \leq t \leq \delta$  and  $x_{\delta}(t) = 0$  otherwise. Then  $x_{\delta} \in S_1$  and it follows from (4.51) that

$$(4.52) \quad \text{m.u.b.} \left| \frac{1}{\delta} \int_s^{s+\delta} J(t)dt \right| = \text{m.u.b.} \left| \frac{1}{\delta} \int_{-s}^{-s+\delta} J(t)dt \right| \leq D.$$

But since  $J(t)$  is integrable over each finite interval,  $(1/\delta) \int_s^{s+\delta} J(t)dt$  is a continuous function of  $s$  for each  $\delta > 0$ . Hence it follows from (4.52) that

$$(4.53) \quad \left| \frac{1}{\delta} \int_s^{s+\delta} J(t)dt \right| \leq D, \quad 0 < \delta < 1, \quad -\infty < s < \infty.$$

But, by one form of the fundamental theorem of the calculus,

$$(4.54) \quad \lim_{\delta \rightarrow 0} \frac{1}{\delta} \int_s^{s+\delta} J(t)dt = J(s)$$

for almost all  $s$ . Hence  $|J(s)| \leq D$  for almost all  $s$  so that

$$(4.55) \quad \beta = \text{m.u.b.} |J(t)| \leq D.$$

Therefore  $J \in B$  and Theorem 4.1 is proved.

To prove Theorem 4.2, let  $J \in B$  so that  $\beta = \text{m.u.b.} |J(t)| < \infty$ . If  $x \in L$ , then for each  $s$

$$\begin{aligned}
 |y(s)| &= \left| \int_{-\infty}^{\infty} J(t)x(s+t)dt \right| \leq \int_{-\infty}^{\infty} |J(t)| |x(s+t)| dt \\
 &\leq \beta \int_{-\infty}^{\infty} |x(s+t)| dt = \beta \int_{-\infty}^{\infty} |x(t)| dt
 \end{aligned}$$

and (4.21) follows. If it be assumed that

$$(4.56) \quad \text{l.u.b.}_{-\infty < s < \infty} |y(s)| \leq C \int_{-\infty}^{\infty} |x(t)| dt, \quad x \in L,$$

where

$$(4.57) \quad C < \beta = \text{m.u.b.}_{-\infty < s < \infty} |J(t)|,$$

then we can set  $D=C$  in (4.51) to obtain  $D=C$  in (4.55) and have a contradiction of (4.57). Therefore if  $C < \beta$ , then  $x \in L$  exists for which (4.23) holds, and proof of Theorem 4.2 is complete.

**5. Conditions for continuity of  $y(s)$ .** The following theorem, which we give mainly for comparison with other theorems, is easily proved.

**THEOREM 5.1.** *In order that  $J(t)$  may be such that*

$$(5.11) \quad y(s) = \int_{-\infty}^{\infty} J(t)x(s+t)dt$$

*exists for all real  $s$  and is continuous whenever  $x \in L$ , it is necessary and sufficient that  $J(t)$  be measurable and essentially bounded.*

Necessity is a consequence of Theorem 2.5. If  $\beta = \text{m.u.b. } |J(t)| < \infty$  and  $x \in L$ , then the estimate

$$\begin{aligned}
 |y(s+h) - y(s)| &\leq \int_{-\infty}^{\infty} |J(t)| |x(s+h+t) - x(s+t)| dt \\
 &\leq \beta \int_{-\infty}^{\infty} |x(s+h+t) - x(s+t)| dt \\
 &= \beta \int_{-\infty}^{\infty} |x(t+h) - x(t)| dt,
 \end{aligned}$$

together with the fact that  $x \in L$  implies that the last member converges to 0 as  $h \rightarrow 0$ , shows that  $y(s)$  is uniformly continuous.

It is interesting to note in connection with Theorem 5.1 and earlier theorems that the hypothesis that  $y(s)$  exists and is continuous for all real  $s$  whenever  $x \in S$  does not imply that  $J \in B$ . To prove this, let  $J(t)$  be a function in  $L$  which is not essentially bounded and which vanishes outside some finite in-

terval  $a \leq t \leq b$ , say  $J(t) = t^{-1/2}$  over  $0 < t < 1$  and  $J(t) = 0$  otherwise. Let  $x \in S$ . Then for each fixed real  $s_0$

$$y(s_0) = \int_{-\infty}^{\infty} J(t)x(s_0 + t)dt$$

exists since  $x(s_0 + t)$  is measurable and bounded over the finite interval  $a \leq t \leq b$  outside of which  $J(t)$  vanishes. If  $K$  is chosen such that  $|x(s_0 + t)| < K$  over  $a - 1 \leq t \leq b + 1$ , then when  $|h| < 1$

$$\begin{aligned} |y(s_0 + h) - y(s_0)| &\leq \int_{-\infty}^{\infty} |J(t - h) - J(t)| |x(s_0 + t)| dt \\ &\leq K \int_{-\infty}^{\infty} |J(t - h) - J(t)| dt, \end{aligned}$$

and, since the last integral converges to 0 with  $h$ ,  $y(s)$  is continuous at  $s_0$ . Thus Theorem 5.1 will fail if the phrase "whenever  $x \in L$ " is replaced by the phrase "whenever  $x \in S$ ."

**6. Some examples.** Theorem 2.1 differs from Theorems 3.1 and 4.1 in that the hypothesis of Theorem 2.1 involves the special class  $S_U$  of unit step functions while the hypotheses of Theorems 3.1 and 4.1 involve the larger class  $S$ . We are going to show that Theorems 3.1 and 4.1 will fail if  $S$  is replaced by  $S_U$  in their statements. For the case of Theorem 4.1, we observe that if  $x \in S_U$  then  $x \in B$  and hence that if  $J \in L$  then  $y \in B$ ; therefore the hypothesis that  $y \in B$  whenever  $x \in S_U$  does not imply that  $J \in B$ . For the case of Theorem 3.1, let  $J(t) = e^{2\pi it}$ . If  $x \in S_U$ , then

$$y(s) = \int_{-\infty}^{\infty} e^{2\pi it} x(s + t) dt$$

exists for each  $s$  since  $x(s + t) \in L$  and  $e^{it}$  is measurable and bounded; and the fact that  $x(s + t)$  is constant over unit intervals, together with the fact that the integral of  $e^{2\pi it}$  over each unit interval is 0, implies that  $y(s) = 0$  and hence  $y \in L$ . Since  $J \notin L$  fails, the hypothesis that  $y \in L$  whenever  $x \in S_U$  does not imply that  $J \in L$ .

We show also that none of Theorems 2.1, 3.1 and 4.1 will hold if the hypotheses are relaxed to require that  $y(s)$  have the stated property only when  $x(t)$  is an ordinary step function. By an *ordinary step function*, we mean a finite linear combination of simple step functions; a *simple step function* being a function  $\xi(t)$  such that  $\xi(t) = 1$  for all  $t$  in the interior of some finite interval  $I$  and  $\xi(t) = 0$  for all  $t$  outside the closure of  $I$ . Except for the inconsequential fact that we do not require ordinary step functions to have right-hand continuity at end points of intervals, an ordinary step function may be described as a function in  $S$  which vanishes outside some finite interval; hence each

ordinary step function is equal, for all except a finite set of values of  $s$ , to a function in  $S$ .

For the case of Theorems 2.1 and 4.1, let

$$(6.11) \quad J_1(t) = te^{it^2}.$$

If  $x(t) = 1$  when  $a < t < b$  and  $x(t) = 0$  when  $t < a$  and when  $t > b$ , and  $y_1(s)$  is the  $J_1$  transform of  $x(t)$ , then

$$(6.12) \quad y_1(s) = \int_{a-s}^{b-s} te^{it^2} dt = [e^{i(b-s)^2} - e^{i(a-s)^2}]/2i$$

so that  $y_1(s)$  exists for all  $s$  and is continuous over  $-\infty < s < \infty$ , and

$$(6.13) \quad |y_1(s)| \leq 1, \quad -\infty < s < \infty.$$

It follows easily that the  $J_1$  transform of each ordinary step function exists for all  $s$  and is bounded and continuous. But  $J_1$  is not essentially bounded, and the condition

$$(6.14) \quad \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+A} |J_1(t)| dt < \infty$$

fails for each  $A > 0$ . Thus the hypothesis that  $y(s)$  exists and is bounded and continuous over  $-\infty < s < \infty$  whenever  $x(t)$  is an ordinary step function implies neither the conclusion of Theorem 4.1 nor the conclusion of Theorem 2.1.

For the case of Theorem 3.1, let

$$(6.21) \quad J_2(t) = e^{it^n}$$

where  $n$  is a fixed real number greater than 2. If  $x(t) = 1$  when  $a < t < b$  and  $x(t) = 0$  when  $t < a$  and when  $t > b$ , and if  $y_2(s)$  is the  $J_2$  transform of  $x(t)$ , then we have

$$(6.22) \quad y_2(s) = \int_{-\infty}^{\infty} J_2(t)x(s+t)dt = \int_{a-s}^{b-s} e^{it^n} dt.$$

Integration by parts gives, when  $|s|$  is so great that the interval  $a-s \leq t \leq b-s$  does not contain the origin,

$$(6.23) \quad y_2(s) = \left[ \frac{t^{1-n}}{in} e^{it^n} \right]_{a-s}^{b-s} - \frac{1-n}{in} \int_{a-s}^{b-s} t^{-n} e^{it^n} dt.$$

Hence, for such values of  $s$ ,

$$(6.24) \quad |y_2(s)| \leq n^{-1} [ |b-s|^{1-n} + |a-s|^{1-n} ] + (b-a) [ |b-s|^{-n} + |a-s|^{-n} ]$$

so that

$$(6.25) \quad \limsup_{|s| \rightarrow \infty} |s|^{n-1} |y_2(s)| \leq 2/n.$$

Since  $y_2(s)$  is continuous,  $n > 2$ , and (6.25) holds, we have  $y_2 \in L$ . Thus the  $J_2$  transform of each simple step function is bounded, continuous, and in  $L$ ; and it follows that the  $J_2$  transform of each ordinary step function also has these properties. But  $J_2 \in L$  fails. This shows that the hypothesis that  $y$  is bounded, continuous, and in  $L$  whenever  $x$  is an ordinary step function does not imply that  $J \in L$ .

In case  $n = 2$ , the transformation determined by the kernel (6.21) becomes

$$(6.26) \quad y(s) = \int_{-\infty}^{\infty} e^{ist} x(s+t) dt = \int_{-\infty}^{\infty} e^{i(t-s)^2} x(t) dt;$$

and this can be written in the form

$$(6.27) \quad \eta(s) = \int_{-\infty}^{\infty} e^{-2ist} \xi(t) dt$$

where

$$(6.28) \quad \eta(s) = y(s)e^{-is^2}, \quad \xi(t) = x(t)e^{it^2}.$$

The function  $\eta(s)$  of (6.27) differs in only a simple way from the Fourier transform of  $\xi(t)$ . If  $\xi(t)$  is a simple step function, it is easy to compute  $\eta(s)$  and to show that  $\eta(s)$  is not in class  $L$ .

**7. The class  $K$  of measurable functions satisfying (2.12).** The classes  $L$  and  $B$ , and the linear vector metric complete spaces associated with them, are well known. (See, for example, the book of Banach previously cited.) In Theorem 2.1 we were led to the class  $K$  of measurable functions, a member of which we now denote by  $x(t)$ , such that

$$(7.1) \quad \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+A} |x(t)| dt < \infty$$

for each  $A > 0$ . The class  $K$  contains all elements of  $L$  and all elements of  $B$ . It is easy to show that the class  $K$  is linear, that is, if  $x_1, x_2 \in K$  and  $c_1, c_2$  are constants, then  $c_1x_1 + c_2x_2 \in K$ . In terms of a fixed  $A > 0$  and a number  $\phi(A) > 0$  let the *norm* of each  $x \in K$  be defined by

$$(7.2) \quad \|x\| = \|x\|_A = \text{l.u.b.}_{-\infty < u < \infty} \phi(A) \int_u^{u+A} |x(t)| dt.$$

Dependence of  $\|x\|$  on  $A$  is illustrated by the fact that if  $x_0(t) = |t|^{-1/2}$ , then  $x_0 \in K$  and

$$(7.3) \quad \|x_0\| = \phi(A) \int_{-A/2}^{A/2} |x_0(t)| dt = (8A)^{1/2} \phi(A).$$

There seems to be no compelling reason why one choice of  $A$  and  $\phi(A)$  should be preferred over another. If  $x \in L$  and  $\phi(A)=1$ , then  $\|x\|_A$  converges to  $\int_{-\infty}^{\infty} |x(t)| dt$  as  $A \rightarrow \infty$ ; if  $x \in B$  and  $\phi(A)=1/A$ , then  $\|x\|_A$  converges to m.u.b.  $|x(t)|$  as  $A \rightarrow 0$ .

Assuming now that  $A > 0$  and  $\phi(A) > 0$  are fixed, it is easy to see that the class  $K$  becomes a linear vector metric space when the distance between two elements  $x_1$  and  $x_2$  of  $K$  is defined by  $\|x_2 - x_1\|$ . We conclude by showing that this space is complete. Let  $x_1, x_2, \dots$  be a Cauchy sequence in  $K$  so that  $\|x_m - x_n\| \rightarrow 0$  as  $m, n \rightarrow \infty$ . Then as  $m, n \rightarrow \infty$

$$(7.4) \quad I_{m,n,r} \equiv \int_{rA}^{(r+1)A} |x_m(t) - x_n(t)| dt$$

converges to 0 uniformly in  $r$ . Since space  $L$  is complete, there is for each integer  $r=0, \pm 1, \pm 2, \dots$  a function  $\xi_r(t)$  defined over  $rA \leq t < (r+1)A$  such that

$$(7.5) \quad \lim_{n \rightarrow \infty} \int_{rA}^{(r+1)A} |\xi_r(t) - x_n(t)| dt = 0.$$

If we let  $\xi(t)$  be the function defined over  $-\infty < t < \infty$  which agrees with  $\xi_r(t)$  in the interval  $rA \leq t < (r+1)A$ , then for each real  $r$

$$(7.6) \quad I_{n,r} \equiv \int_{rA}^{(r+1)A} |\xi(t) - x_n(t)| dt$$

converges to 0 as  $n \rightarrow \infty$ . The inequality

$$(7.7) \quad |I_{m,r} - I_{n,r}| \leq I_{m,n,r},$$

together with the fact that the right member converges to 0 uniformly in  $r$  as  $m, n \rightarrow \infty$  implies that the left member converges to 0 uniformly in  $r$  as  $m, n \rightarrow \infty$  and hence that  $I_{n,r}$  converges uniformly in  $r$  as  $n \rightarrow \infty$ . But  $I_{n,r}$  converges to 0 as  $n \rightarrow \infty$ . Hence  $I_{n,r}$  converges uniformly to 0 as  $n \rightarrow \infty$ , that is,

$$(7.8) \quad \lim_{n \rightarrow \infty} \text{l.u.b.}_{-\infty < r < \infty} \int_{rA}^{(r+1)A} |\xi(t) - x_n(t)| dt = 0,$$

or

$$(7.9) \quad \lim_{n \rightarrow \infty} \text{l.u.b.}_{-\infty < u < \infty} \int_u^{u+A} |\xi(t) - x_n(t)| dt = 0.$$

This implies that  $\xi \in K$ , and on multiplying by the constant  $\phi(A)$  we obtain  $\lim \| \xi - x_n \| = 0$ . Thus each Cauchy sequence  $x_n$  in  $K$  has a limit in  $K$ , and completeness of the space  $K$  is established.

CORNELL UNIVERSITY,  
ITHACA, N. Y.



## TOPOLOGICAL GROUP FOUNDATIONS OF RIGID SPACE GEOMETRY

BY

DEANE MONTGOMERY AND LEO ZIPPIN

Dedicated to the memory of Bella Zippin, mother of one and friend of both of us.

1. Hilbert, after building up geometry from a point of view which relegated continuity considerations to the background [4], built up plane geometry afresh [5] on the foundation of groups of *homeomorphisms* of the *number plane*, both "continuity" concepts. It is this latter point of view with which we are concerned in this paper. Hilbert carried out this program only for the plane but he hinted that it might be possible to carry it out in a somewhat similar way for three-space. Kerékjártó took up the problem [6] for three-space and made a great deal of progress with it, but, as he wrote before the recent developments in topological groups, he found it necessary to employ a stronger set of axioms than is necessary now.

Relying on the theory of topological groups we recently characterized the rotation group of three-space [9], and in commenting on that work P. A. Smith [12] suggested that it might provide the basis for an extension of Hilbert's program to three-space.

The purpose of this paper is twofold. In the first place we shall characterize the classical space geometries on the basis of a fairly weak set of axioms, and in the second place we shall show that Hilbert's axioms for the case of the plane can be weakened by replacing what might be called his "three-point condition" by a two-point condition. We achieve this latter purpose more or less incidentally to the first.

In comparing our set of three-space axioms with Hilbert's axioms for the plane we find that the first axiom is the same in both cases. The third axiom of this paper is weaker than Hilbert's, and our second purpose above is to show that this weaker axiom also suffices in Hilbert's case. Our second axiom is weaker than Hilbert's second axiom in that it relates to the subgroup leaving a single point fixed, but it is incomparably stronger in what it asks of that one subgroup.

We do not, in this paper, settle the question of whether or not Hilbert's second axiom is adequate for three-space geometries. This question is bound up with an unsolved problem concerning transformation groups.

Finally we wish to remark that instead of assuming that the space we are dealing with is ordinary three-space, it is only necessary to make certain topological assumptions on the space from which it follows by virtue of the

Presented to the Society, October 28, 1939; received by the editors December 20, 1939.

same group axioms that the space is actually a number-space. But we reserve discussion of this matter for another occasion. In this connection see the abstract by the authors in the Bulletin of the American Mathematical Society, vol. 45 (1939) (no. 349).

**2. The axioms.** We formulate two sets of axioms, the first set for the plane, and the second set for three-space. The first set, to which we proceed immediately, is the set used by Hilbert except that it has been materially weakened in the manner described.

We assume then that there is given a system  $(E_2, G)$  where  $E_2$  is the number-plane and  $G$  is a set of sense preserving homeomorphisms of this plane, and that this set satisfies the following conditions:

**2.1. The system  $G$  is a group.**

The assumption tacitly implicit in the above is that  $G$  is effective, that is, that no element except the identity leaves all of  $E_2$  fixed.

With each point  $x$  in space there exists a subgroup  $G_x$  consisting of elements of  $G$  which leave  $x$  fixed.

**2.2. If  $x$  is any point of  $E_2$  and  $y$  is distinct from  $x$ , then  $G_x(y)$  is infinite.**

This axiom could be reformulated so that it would be entirely analogous to our axiom 2.2' for three-space but we do not carry this out. It would involve almost no change in the work.

**2.3. Let  $(x, y)$  and  $(x', y')$  be two pairs of points of  $E_2$ , where the points of a pair are not necessarily distinct. If there exist pairs  $(x_n, y_n)$  and  $(x'_n, y'_n)$ , the first arbitrarily near  $(x, y)$ , the second arbitrarily near to  $(x', y')$ , and if there exists an element of  $G$  taking  $(x_n, y_n)$  to  $(x'_n, y'_n)$ , then there exists an element of  $G$  taking  $(x, y)$  to  $(x', y')$ .**

As we have said, this set of axioms is exactly Hilbert's except that the third axiom has been weakened to a condition on pairs instead of triples of points.

We now formulate our axioms for three-space. We assume that there is given a system  $(E_3, G)$  where  $E_3$  is ordinary three-space and  $G$  is a set of sense preserving homeomorphisms of  $E_3$  satisfying the following conditions:

**2.1'. The same as 2.1.**

**2.2'. There exists a point  $p$  of  $E$  such that the group  $G_p$  is a proper subgroup of  $G$  and for a sequence of points  $p_n$  approaching  $p$  the sets  $G_p(p_n)$  are at least two dimensional.**

**2.3'. The same as 2.3.**

Occasionally we shall refer to the situation described by the first set of axioms as the *plane case* and to the situation described by the second set as



the *space case*. In both cases we shall prove that ordinary geometric concepts such as "line" and "distance" (and in the space case "plane") may be defined in terms of  $G$  in such a way that we obtain either euclidean or hyperbolic geometry and that  $G$  is the group of rigid motions of the corresponding geometry we obtain. We do this in the plane case by proving Hilbert's axioms [5], that is, by proving that the two-point condition 2.2 implies the three-point condition. The space case we treat in detail and show in detail that there are only the two systems.

Our approach to this problem differs from Hilbert's in one important respect. Hilbert analyzes more or less directly the topological nature of the orbit  $G_p(x)$ . In three-space this course seems to us not feasible until much more is known about strongly homogeneous subsets of space. But even granting such knowledge our procedure has the advantage of making available the results of topological group theory. Thus, we proceed at once to a study of the group  $G_p$  as a topological transformation group. In brief summary, we first confine our attention to a suitable invariant neighborhood of the point  $p$  where orbits under  $G_p$  can be proved compact. We form the effective group in this neighborhood, and show that any sequence of elements of this group has a subsequence which converges to a homeomorphism of the neighborhood into itself. We then augment our group by the addition of such homeomorphisms. It transpires, only considerably later, that this enlargement is an illusory one. The enlarged group is then shown to be a compact topological transformation group on a "locally euclidean" space. From our previous work, we then know that our orbits are necessarily manifolds, and it is not difficult to show that they are indeed spheres. From an earlier paper of ours we learn also the complete structure of the group and its behavior in the neighborhood. We are now in a position to show, by an argument patterned on one of Hilbert's, that the neighborhood above coincides with space.

The use of a "two-point" rather than "three-point" axiom shows itself in one or two interesting ways in the study of  $G_p$  but becomes a matter of considerable moment in the study of the group  $G$  as a topological transformation group of the space. This argument is given in §12. To this point the case of the plane or of space may be treated more or less simultaneously, and it seems to us, in fact, that much of this generalizes with no great difficulty to four-space and perhaps farther.

The remainder of the paper is devoted to a study of the geometry induced in space by the group  $G$ . Here, after a few paragraphs, we are on ground already explored by Kerékjártó. His paper was not known to us until our own had been completed, and we carry out the program essentially as we had it. We do this in part for completeness sake, in part because the form in which our solution is set differs sufficiently from Kerékjártó's. At one point in proving the linearity of our planes we borrow a very ingenious idea from Kerékjártó's paper which shortens considerably an argument of our own.

3. In a great part of the paper we treat the two cases simultaneously, calling the space (which of course is either  $E_2$  or  $E_3$ ) simply  $E$ . When we speak of a sphere or a rotation group we of course mean the one appropriate to the dimension of the space.

Let  $p$  be a point of  $E$ , which for  $E_2$  may be any point, but which for  $E_3$  is to be the point specified by 2.2'.

Hilbert points out that for any  $x$  the set  $G_p(x)$  is closed. This is a consequence of 2.3. Thus: let  $x_n$  be a sequence of points in  $G_p(x)$  converging to a limit point  $x_0$ . There are elements  $g_n$  in  $G_p$  such that  $x_n = g_n(x)$ . The element  $g_n$  takes the pair  $(p, x)$  to the pair  $(p, x_n)$ . By 2.3 there is a  $g$  in  $G$  which takes  $(p, x)$  to  $(p, x_0)$  and this element certainly is in  $G_p$ .

By a similar method it is seen that  $G(x)$ , which ultimately will be shown to coincide with  $E$ , is closed.

4. Let  $R$  be the set of points  $x$  such that  $G_p(x)$  is compact. This set is not vacuous for it contains  $p$ .

LEMMA 1. *The set  $R$  is open.*

Let  $x$  be any point in  $R$ , and let  $B$  be a conditionally compact open set containing  $G_p(x)$  in its interior. Let  $S$  be the boundary of  $B$ . We assert that there is an open set  $V$  containing  $G_p(x)$  such that  $G_p$  carries no point of  $S$  inside  $V$ . Otherwise there must exist a set of elements  $g_n$  in  $G_p$  and a set of points  $b_n$  in  $S$  such that at least one point of the set  $g_n(b_n)$  lies in every neighborhood of the compact set  $G_p(x)$ . There is no loss in assuming that  $b_n$  approaches some point  $b$  in  $S$  and  $g_n(b_n)$  approaches a point  $a$  in  $G_p(x)$ . But then there must be an element of  $G_p$  taking  $b$  to  $a$  which means that a point of  $S$  is in  $G_p(x)$  contrary to the choice of  $B$ .

Let  $W$  denote those components of the above determined  $V$  which include points of  $G_p(x)$ . The set  $W$  is open and no element of  $G_p$  carries a point of  $W$  outside of  $B$ . For such an element would also leave an element of  $W$  inside  $B$  (any point namely in which  $W$  meets  $G_p(x)$ ), and it would therefore carry some point of  $W$  into  $S$  which is impossible.

It has now been shown that all points of  $W$  have orbits inside  $B$ . Therefore every point of  $W$  has a compact orbit and  $x$  is in an open set  $W$  all of whose points have compact orbits as the lemma demands.

4.1. The proof shows even more than is required in Lemma 1. It shows, for any point  $x$  in  $R$ , that  $x$  is an interior point of a set  $W$ , such that  $G_p(W)$  is a conditionally compact set. If  $\bar{W}$  is the closure of  $W$ , then  $G_p(\bar{W})$  is compact and  $x$  is seen to be an interior point of a set  $\bar{W}$  such that  $G_p(\bar{W})$  is compact. These facts together with the Heine-Borel theorem enable us to state the following lemma.

LEMMA 2. *If  $M$  is any compact subset of  $R$ , then  $G_p(M)$  is compact.*

5. We consider now the action of  $G_p$  on  $R$ . Conceivably  $G_p$  has a non-

trivial subgroup leaving all of  $R$  fixed. Later this possibility will be ruled out, but meanwhile we must take account of it. Let  $G_p^1$  be the subgroup leaving all of  $\bar{R}$  fixed. This subgroup includes the identity at any rate, and  $G_p/G_p^1$ , which will be denoted for brevity by  $H$ , is an effective transformation group of  $\bar{R}$ .

Our next task is to show that  $H$  may be extended to become a compact transformation group of  $\bar{R}$ . It will be assumed that  $E$  is assigned a bounded metric, say the metric of a three-sphere or a two-sphere according to the case, which is obtained by adding a point to  $E$ . This means that we can define a distance between any two transformations of  $E$  into itself or between any two transformations of a subset of  $E$  into itself. For example if  $f$  and  $g$  are two transformations of  $\bar{R}$  into itself

$$d(f, g) = \text{l.u.b. } d[f(x), g(x)]$$

where  $x$  ranges over  $\bar{R}$ . Under this distance  $H$  becomes a metric space.

5.1. LEMMA 3. *If a sequence of elements of  $G$  converges everywhere on a set  $M$  to a limit  $h$ , then  $h$  is continuous on  $M$ . If  $M$  is compact the convergence is uniform.*

Let us prove first that  $h$  is continuous. Let  $m$  be any point of  $M$  and let  $S$  be any sphere with  $f(m)$  as a center. Let  $m_i$  be a sequence of points of  $M$  approaching  $m$ . We shall show that almost all the points  $h(m_i)$  are inside or on  $S$ . Assume that this is not true for an infinite subsequence, say  $m_{k_i}$ , and let  $m_{k_i}m$  be a short arc joining  $m_{k_i}$  to  $m$ . Since  $g_{n_i}$  approaches  $h$  and since  $h(m_{k_i})$  is outside  $S$  by assumption, there will certainly be an integer, say  $n_i$ , such that  $g_{n_i}(m_{k_i})$  is outside  $S$ . We may assume without loss of generality that every  $g_{n_i}(m)$  is inside  $S$ . Hence there is a point, say  $y_{k_i}$ , on the arc  $m_{k_i}m$  such that  $g_{n_i}(y_{k_i})$  is on  $S$ . Assume that  $g_{n_i}(y_{k_i})$  approaches a point  $b$  on  $S$ . Then  $g_{n_i}$  takes the pair  $(y_{k_i}, m)$  which is near  $(m, m)$  to the pair  $[g_{n_i}(y_{k_i}), g_{n_i}(m)]$  which is near  $[b, h(m)]$ . This is a contradiction from which the continuity of  $h$  follows.

We shall next show that the convergence is uniform in case  $M$  is compact. If the convergence is not uniform, there is for some positive number  $\epsilon$  an infinite set of indices  $k_1, k_2, \dots$  and a set of points  $m_1, m_2, \dots$  in the set  $M$  such that

$$d[g_{k_i}(m_i), h(m_i)] \geq \epsilon.$$

There is no loss in assuming that the sequence  $m_i$  converges to  $m$  and that  $g_{k_i}(m_i)$  converges to a point  $b$ . From the above inequality  $b$  and  $h(m)$ , which is the limit of  $h(m_i)$  by the continuity, are distinct. The transformations  $g_{k_i}$  therefore take the pair  $(m_i, m)$  which is near  $(m, m)$  to the pair  $[g_{k_i}(m_i), g_{k_i}(m)]$  which is near  $[b, h(m)]$ . By 2.3 there must be an element of  $G$  taking  $m$  to each of the distinct points  $b$  and  $h(m)$ . This contradiction shows that the convergence is uniform.

LEMMA 3.1. *Let  $g_n$  be a sequence of elements of  $G$  converging to  $h$  everywhere on a set  $M$ . Then if  $m_i$  approaches  $m$  it follows that  $g_i(m_i)$  approaches  $h(m)$ .*

The set  $B$  containing the points  $m$  and all  $m_i$ 's is compact. Hence on this set  $g_n$  converges uniformly to  $h$ . Let  $\epsilon$  be any positive number. For all  $n$  greater than an integer  $N_1$ ,

$$d[g_n(x), h(x)] < \epsilon/2$$

for all  $x$  in  $B$ . Since  $h$  is continuous, there will be an integer  $N_2$  such that if  $n$  is greater than  $N_2$

$$d[h(m_n), h(m)] < \epsilon/2.$$

For all  $n$  greater than  $N_1$  and  $N_2$  we have not only this latter inequality but we have as a consequence of the first inequality

$$d[g_n(m_n), h(m_n)] < \epsilon/2.$$

The last two inequalities yield the desired conclusion.

LEMMA 3.2. *If a sequence  $f_n$  of elements of  $G_p$  converges everywhere on a compact set  $M$ , invariant under  $G_p$ , to a transformation  $f$ , then  $f$  is a homeomorphism of  $M$  into itself.*

In view of Lemma 3 it is only necessary to show that  $f$  has a single valued inverse and that  $f(M) = \bar{M}$ .

If  $f$  does not have a single valued inverse, there must be two distinct points  $b$  and  $c$  such that  $f(b) = f(c)$ . Then the elements  $f_n$  take the pair  $(b, c)$  to the pair  $[f_n(b), f_n(c)]$  and by 2.3 there is an element in  $G$  which takes both  $b$  and  $c$  to  $f(b)$ . This contradiction shows that  $f$  is one-one, and since  $M$  is compact  $f$  must take  $M$  homeomorphically to  $f(M)$ . We know that  $f(M)$  is a subset of  $M$ , and to complete the proof of the lemma it remains only to show that  $f(M)$  coincides with  $M$ .

Let  $b$  be any point in  $M$ . Then there is an element  $m_n$  in  $M$  such that  $f_n(m_n) = b$ . Assume that the sequence  $m_n$  approaches a point  $m$ . Now  $f_n$  takes the pair  $(m_n, m)$  which is near  $(m, m)$  to the pair  $[b, f_n(m)]$  which is near  $[b, f(m)]$ . Hence some element of  $G$  takes  $m$  to both points  $b$  and  $f(m)$  which is possible only if  $f(m) = b$ .

5.2. LEMMA 4. *The group-space  $H$  defined in §5 is conditionally compact.*

Yet  $Y$  be a countable dense subset of  $R$ . Let  $g_n$  be an infinite sequence of elements of  $H$ . Strictly speaking the  $g_n$ 's are not elements of  $G_p$  but there are elements of  $G_p$  coinciding with these elements on  $\bar{R}$ , and properties of the group  $G_p$  may be used in examining the  $g_n$ 's.

For any point  $y$  in  $Y$  the sequence  $g_n(y)$  is conditionally compact and has a convergent subsequence. The limit of this sequence belongs to  $G_p(y)$  and is a

point of  $R$ . By the diagonal process there exists a subsequence  $f_n$  of the elements  $g_n$  such that  $f_n(y)$  converges to a unique point of  $R$  for every element  $y$  of  $Y$ . Then, on  $Y$ , the sequence  $f_n$  converges to a pointwise limit function  $f$ .

It will now be shown that  $f$  is uniformly continuous in every conditionally compact open subset  $R_1$  of  $R$ . In order to do this it must be shown that for every positive  $\epsilon$  there exists a positive  $d$  such that whenever  $y$  and  $y'$  of  $Y \cdot R_1$  are nearer to each other than  $d$ , the corresponding  $f(y)$  and  $f(y')$  are nearer than  $\epsilon$ . If this is not the case, there must exist in  $Y$  a sequence of points  $y_n$  and  $y'_n$  which may be supposed to converge to the same point  $z$  of  $\bar{R}_1$  such that  $f(y_n)$  and  $f(y'_n)$  also converge and converge to two distinct points  $y$  and  $y'$  in  $\bar{R}_1$ . This means that some of the elements  $g_n$  take a pair of points near  $z$  to a pair near  $(y, y')$ . By 2.3 there is an element of  $G$  carrying the point  $z$  to the two points  $y$  and  $y'$ . This is manifestly impossible and the contradiction establishes the uniform continuity of  $f$  on the set  $R_1 \cdot Y$ .

This uniform continuity of  $f$  permits us to extend it, and we assume it is so extended, to a single valued continuous transformation of  $R$  (which of course is locally compact) into itself. The sequence  $f_n$ , which originally was known to converge only on  $Y$ , is now seen to converge everywhere on  $R$  to the transformation  $f$ . By Lemma 3 the convergence is uniform on compact sets which implies that the sequence  $f_n$  of elements of  $H$  must be a Cauchy sequence because our metric brings two functions close which agree closely outside of a neighborhood of "infinity."

5.3. When we speak of a topological transformation group we use the term with the definition as given in our papers referred to in the bibliography.

LEMMA 5. *The group  $H$  may be extended to a compact group  $\bar{H}$  which is an effective topological transformation group of  $R$ .*

The space  $H$  is conditionally compact so that if the space is made complete the resulting space  $\bar{H}$  will be compact. The preceding lemmas and their proofs show that Cauchy sequences of  $H$  will actually converge to homeomorphisms of  $R$  into itself. The space  $\bar{H}$  is therefore a transformation group of  $R$ . To be sure that it is a topological transformation group we must prove that if  $g_n$  approaches  $g$ , these being elements in  $\bar{H}$ , and  $x_n$  approaches  $x$  in  $R$  then  $g_n(x_n)$  approaches  $g(x)$ . This follows as in the proof of Lemma 3.1.

That  $\bar{H}$  is an effective group follows from the fact that distinct elements of the space  $\bar{H}$  arise from nonequivalent Cauchy sequences and these give rise to distinct limiting transformations.

6. The present section contains some simple, purely group theoretical, considerations which will be of use to us later.

THEOREM 1. *The only two dimensional manifolds which are coset spaces (orbits) of a compact connected group  $H$  are the two-sphere, torus, and projective plane.*



There is no loss in assuming that  $H$  is effective in its action on the manifold so that  $H$  is a Lie group [10]. The theorem then follows from the corresponding theorem on Lie groups due to Cartan [2].

We note without giving the proof, which is not difficult, the following: If a circle group operates on a torus and has a fixed point, then it must leave every point of the torus fixed.

**THEOREM 2.** *The only compact group of sense preserving transformations which can act effectively and transitively on a two-sphere  $M$  is the group of rigid rotations of the two-sphere.*

For connected groups the theorem is true [9]. If  $H$  is the group, let  $H^*$  be the identity component of  $H$ , and let  $x$  be a point of the sphere. The dimension of  $H^*(x)$  is the same as the dimension of  $H(x)$ , namely two, and therefore  $H^*(x)$  coincides with  $M$ . Since  $H^*$  is connected it must be the two-sphere rotation group. Assuming  $H^*$  is not all of  $H$  means that, for some  $y$ ,  $(H^*)_y$  is a proper subgroup of  $H_y$  because every element of  $H$ , being sense preserving, has a fixed point; the connectedness of the group  $(H^*)_y$  (it is of course circular) shows that it is the component of the identity of  $H_y$ . Then  $(H^*)_y$  is invariant in  $H_y$ . Let  $M^*$  be the decomposition space of  $M$  under  $(H^*)_y$ . The group  $H_y/(H^*)_y$  acts on this space which is an interval. Hence  $H_y/(H^*)_y$  contains only the identity, or it is a group of two elements which merely interchanges the ends of  $M^*$  while leaving a "middle" point fixed. The latter possibility cannot occur, for if it did  $H_y$  would contain an element moving  $y$ . Hence  $H_y/(H^*)_y$  contains only the identity element and  $(H^*)_y$  is not a proper subgroup of  $H_y$  as we assumed. The contradiction shows that  $H^*$  coincides with  $H$  and that  $H$  is the group of rigid rotations of the two-sphere.

**COROLLARY.** *The only compact group of sense preserving transformations of three-space into itself with at least one two dimensional orbit is the two-sphere rotation group.*

7. In the present section we confine our attention to the space case. The groups  $H$  and  $G_p$  have the same orbits in  $R$ , and since these orbits are closed  $\bar{H}$  has the same orbits as do  $H$  and  $G_p$ . Let  $H^*$  be the component of the identity of  $\bar{H}$ . The orbit of a point under  $H^*$  will have the same dimension as the orbit of the point under  $\bar{H}$ . Hence  $H^*$  has a sequence of two dimensional orbits approaching the fixed point  $p$ . We will now consider the action of the effective compact transformation group  $H^*$  on the connected locally "euclidean" space  $R^0$ , where  $R^0$  denotes that component of  $R$  which contains  $p$ . (It is conceivable that some subgroup of  $H$  should leave all of  $R^0$  fixed. We assume without changing our notation that this is not the case. There is no loss in this process as we might as well have assumed we were working with  $R^0$  before.)

It follows from theorems on the structure of coset spaces [10] that any two dimensional orbit of  $H^*$  in  $R^0$  must be a two dimensional manifold. Any

two dimensional orbit of  $H^*$  in  $R^0$  must be, therefore, one of three types of manifold, the two-sphere, the projective plane, or the torus. The projective plane cannot be imbedded in  $E$ , so that the number of possibilities is reduced immediately to two, the sphere and torus.

In the orbit space associated with  $R^0$ , call it  $R^*$ , every two dimensional orbit is a cut point of order two precisely. The set of such orbits is open [10]. The space  $R^*$  has one non-cut point, at least, namely the orbit consisting of the point  $p$  only. By the cyclic element theory,  $R^*$  must be either a line, a ray, or an interval. It cannot be a line because it contains a non-cut point. It cannot be an interval for this would mean that  $R^0$ , an open subset of  $E$ , would be compact. Hence  $R^*$  is a ray, and this shows that *all orbits* in  $R^0$ , with the exception of  $p$ , are two dimensional orbits which are *either spheres or tori*. The group  $H$  can be seen to be a Lie group because it operates on a locally euclidean connected space with locally connected orbits [10]. It will now be of dimension three at most [10], and it will be effective on each one of its two dimensional orbits. For, if a subgroup left all of a manifold orbit fixed this same subgroup would leave the whole space fixed by a simple application, as in an earlier paper of ours [8], of a theorem of Newman. If all two dimensional orbits are spheres, then  $H^*$  is the rotation group of a sphere, for this is the only connected compact effective transformation group of a sphere. If all two dimensional orbits are tori, then  $H^*$  is a two dimensional toral group, for this is the only Lie group which can be effective on a torus.

It is intuitively clear that all orbits must be spheres and we now give the proof. We will show that if one orbit is toral then all orbits are. Assume that one orbit  $H^*(x)$  is toral. If  $H^*$  is two dimensional, then  $H^*$  is a toral group and all orbits are tori. If  $H^*$  is three dimensional there must be a circular subgroup  $K$  leaving  $x$  fixed. But if a circular subgroup leaves one point of a torus fixed it must leave every point fixed. Hence  $K$  leaves all of  $H^*(x)$  fixed and, since this separates space, we see by a familiar device that  $K$  leaves all of space fixed. In this case  $H^*$  is not effective. We are therefore led to conclude that all two dimensional orbits are tori.

This last situation is impossible. For if  $H^*$  is a toral group we can form a true section  $B$  of the space, that is, we can find a closed set  $B$  which has one and only one point on each orbit in  $R$ . The set  $B$  will have to be homeomorphic to  $R^*$  and will be a ray. Now let  $H^*(x)$  be a toral orbit inside a neighborhood  $U$  of  $p$  which is homeomorphic to three-space. There will be in  $U - H^*(x)$  a one-cycle  $Z$  which does not bound in  $U - H^*(x)$  and which is outside  $H^*(x)$ ; that is, it is contained in the component of  $U - H^*(x)$  which does not contain  $p$ . Then using the true section  $B$  we may deform  $H^*(x)$  to the point  $p$ . The cycle  $Z$  certainly bounds in  $U - p$ , contradicting its choice.

Therefore not every orbit is a toral orbit and  $H^*$  must be the rotation group of three-space, and every two dimensional orbit must be a two-sphere. Let  $K$  be a circular subgroup of  $H^*$ . There will have to be precisely two points

on each two-sphere orbit left fixed by  $K$ . These two points will define for us a double valued function everywhere on  $R^*$ . The end point of the ray  $R^*$  is an exception when the function is single valued. But at any rate it is possible to pick out of these functional values a true section  $B$  of the entire space. The existence of the ray  $B$  proves  $R^0$  homeomorphic to euclidean three space.

From Theorem 2 of §6 we see that  $\bar{H}$  and  $H^*$  must coincide, but we can conclude even more.

**THEOREM 3.** *The group  $H$  coincides with  $\bar{H}$  and is therefore the rotation group of three-space.*

Let  $x$  be any point of  $R$  distinct from  $p$ . The set  $\bar{H}(x)$  is a two-sphere, and  $H(x)$  coincides with  $\bar{H}(x)$ ; or in other words  $H$ , a subgroup of  $\bar{H}$ , is transitive on the two-sphere. This is possible only if  $H$  is all of  $\bar{H}$  [11].

8. In this section we turn to the plane case, falling back on Lemma 5 where we left it.

**THEOREM 3.1.** *The group  $H$  coincides with  $\bar{H}$  and is a circle group, the rotation group of the plane. The set  $R^0$  is homeomorphic to a plane.*

The group  $\bar{H}$  cannot be totally disconnected, for such a group cannot operate effectively on a locally planar space (as we have shown [8]). Let  $H^*$  be the identity-component of  $\bar{H}$ . As in the space case, the orbits under  $H^*$  must be manifolds and therefore simple closed curves: they are obviously one dimensional. It follows that  $H^*$  is a Lie group and in particular the circle group [10]. As in the space case the group  $\bar{H}$  must operate upon the decomposition space of  $R^0$  by orbits under  $H^*$ : this space is a ray, with  $p$  as end point, and  $H^*$  must be the identity upon it.

Now let  $g$  be some element of  $\bar{H}$ , not the identity. There must be a point  $x$  of  $R^0$  such that  $gx$  is not  $x$ . On the other hand,  $gx$  is a point of  $H^*(x)$  so that for some  $g'$  of  $H^*$  we have

$$g'g^{-1}x = x.$$

Since this element is sense preserving and leaves one point of the circle  $H^*(x)$  fixed, it leaves all  $H^*(x)$  fixed and then all of  $R^0$ . Therefore it must be the identity and we conclude that  $g$  belongs to  $H^*$ . This shows that  $\bar{H}$  coincides with  $H^*$  and is a circle group. Since  $H$  is transitive on  $\bar{H}(x)$ , it is obvious that  $H$  must coincide with  $\bar{H}$  which is effective on  $\bar{H}(x)$ .

Now  $R^0$  is a connected open subset of the plane filled out by a continuous family of simple closed curves and it is intuitively obvious and sufficiently well known that  $R^0$  must be homeomorphic to the plane.

We turn now to a simultaneous consideration of the two cases. What has been done above is summed up in the following theorem.

**THEOREM 4.** *The group  $H$  may be so topologized that it becomes the ordinary*



rotation group of space, and it acts on  $R^0$ , which is homeomorphic to space, as the ordinary rotation group does—in a properly chosen coordinate system.

9. THEOREM 5. *The set  $R^0$  is closed and so coincides with  $E$ .*

The assumption that  $R^0$  is not closed implies that there is an arc  $px$  which is contained in  $R^0$  except for its end point which is not in  $R^0$ . Since  $x$  is not in  $R^0$  (and not in  $R$ ) there is a sequence of elements  $g_n$  of  $G_p$  such that  $g_n(x)$  tends toward infinity. Let  $G_p^{R^0}$  denote all elements of  $G_p$  leaving all of  $R^0$  fixed. Under the homeomorphism taking  $G_p$  to  $G_p/G_p^{R^0} = H$  suppose that  $g_n$  goes to  $\bar{g}_n$ ; assume that the sequence  $\bar{g}_n$  converges to  $\bar{g}$  and that  $\bar{g}$  is the image of an element  $g$  under this homeomorphism.

Let  $S_1$  and  $S_2$  be two spheres each containing  $g(px)$  and such that  $S_2$  lies inside  $S_1$ . We may assume that all points  $g_n(x)$  are outside  $S_1$ . On the arc  $g_n(px)$  there are two points  $g_n(x_n)$  and  $g_n(y_n)$ , where  $x_n$  and  $y_n$  are points of  $px$ , such that  $g_n(x_n)$  is the first point of  $g_n(px)$  on  $S_2$  and  $g_n(y_n)$  is the first point of  $g_n(px)$  on  $S_1$ . The points  $x_n$  and  $y_n$  lie on  $px$  in the order  $px_n y_n x$ .

There is no loss in assuming that the sequences  $x_n$ ,  $y_n$ ,  $g_n(x_n)$ , and  $g_n(y_n)$  converge respectively to  $x'$ ,  $y'$ ,  $x^*$  and  $y^*$ .

We wish to prove that  $x'$  is identical with  $x$ . If it is not,  $x'$  must be a point of  $R^0$  and the points  $x_n$  may also be taken to be in  $R^0$ . Hence  $g_n(x_n) = \bar{g}_n(x_n)$  converges to  $\bar{g}(x')$  which is impossible because  $g(x')$  is inside  $S_2$ . Hence  $x'$  and  $x$  are identical.

This shows that the points  $x_n$  converge to  $x$  and of course the points  $y_n$  must also converge to  $x$ . There are, consequently, elements of  $G$  which take the pair  $(x_n, y_n)$  which is near  $(x, x)$  to the pair  $[g_n(x_n), g_n(y_n)]$  near to the distinct pair  $(x^*, y^*)$ . There is then an element of  $G$  taking  $x$  to the two distinct points  $x^*$  and  $y^*$ . From this contradiction the theorem follows.

COROLLARY. *The group  $G_p^{R^0} = H$ , idle on all of  $R^0$ , contains only the identity and the group  $G_p$  is itself the rotation group of space.*

10. THEOREM 6. *The group  $G$  is transitive on the space  $E$ .*

It has already been remarked that  $G(x)$  is closed for any  $x$  in  $E$ . In particular  $G(p)$  is closed, and in order to prove our theorem it suffices to prove that  $G(p)$  is open.

We know in the plane case as well as in the space case that  $G_p$  is a proper subgroup of  $G$  and there must be an element  $g$  in  $G$  and a point  $q$  distinct from  $p$  such that  $g(p) = q$ . The set  $G_p(q)$  is a sphere. Let  $q'$  denote another point of  $G_p(q)$  and let  $t$  be a varying element of  $G_p$  which takes  $q$  continuously to  $q'$ . There is a neighborhood  $U$  of  $q$  so small that its translation  $U'$  to a neighborhood of  $q'$  has no point in common with its original position. The set  $U'$  is the image of  $U$  under the terminal element of the parameter  $t$ .

There is a neighborhood  $V$  of  $p$  so small that  $g(V)$  is inside  $U$ . As  $q$  is

swept continuously to  $q'$  it must come in contact with the  $g$ -image of every sphere about  $p$  and within  $V$ .

Now let  $s$  be any point of  $V$  and let  $S = G_p(s)$  be its sphere orbit under  $G_p$ . There is a  $t$  such that  $t(q)$  is in  $g(S)$ , that is,

$$t\{g(p)\} = t(q) = gg'(s)$$

for some  $g'$  in  $G_p$ . Then

$$g'^{-1}g^{-1}tg(p) = s.$$

In other words  $p$  may be carried to any element of  $V$  by some element in  $G$ . Therefore  $p$  is an interior point of  $G(p)$  and every point of  $G(p)$  must be an interior point.

11. The fact that  $G$  is transitive on  $E$  tells us the nature of  $G_x$  when  $x$  is distinct from  $p$ . Let  $x = g(p)$ . Then  $G_x = gG_pg^{-1}$  and  $G_x$  is also the rotation group and in a proper system of coordinates "centered" on the point  $x$  it acts as the rotation group ordinarily does.

12. Before such geometric concepts as lines and planes can be studied, we need to analyze the nature of  $G$  as a topological transformation group. It is in this that we encounter the principal difficulties implicit in our use of the "two-point" form rather than Hilbert's "three-point" axiom. In much of this work we shall continue to treat the plane and space cases together.

LEMMA 6. *Let  $x_n$  be a sequence of points converging to  $x$  and let  $g_n$  in  $G$  be such that  $g_n(x_n)$  approaches  $x$ . Then for any  $z$  in  $E$ , the set  $g_n(z)$  is bounded.*

Let  $O$  be the interior of an orbit  $S$  of  $G_x$  which is so large that it surrounds  $z$ , all of the  $x_n$ 's, and all of the points  $g_n(x_n)$ . Let  $O^*$  be the interior of a larger orbit  $S^*$  so that  $S$  is interior to  $O^*$ . Let  $zx_n$  be arcs of  $O$ , and suppose now that for infinitely many of the elements  $g_n$  it is true that  $g_n(z)$  is outside or on  $S^*$ . For these  $n$ 's, and we take it now that all  $n$ 's are such, the arcs  $g_n(x_n z)$  have one point in  $O^*$  and one point not in  $O^*$ . They therefore have a first point  $z_n'$  on  $S^*$ . There is a point  $x_n'$  on  $x_n z$  such that  $z_n' = g_n(x_n')$ . We may suppose that  $x_n'$  converges to a point  $x'$  which is in  $O$  or  $S$ , while  $z_n'$  converges to a point  $z'$  on  $S^*$ . There must be an element  $g$  in  $G$  which leaves  $x$  fixed and carries  $x'$  to  $z'$ . This is by 2.3 because  $g_n$  takes the pair  $(x_n, x_n')$  to the pair  $[g_n(x_n), z_n']$ . However such an element  $g$  is in  $G_x$  and hence leaves  $S$  and its interior invariant so that a contradiction has been reached. Hence  $g_n(z)$  is compact because almost all its elements are inside  $S^*$ .

The orbit  $S^*$  was subject only to the requirement that it surround  $S$  so that the following corollary is true.

COROLLARY. *Let  $x_n$  be a sequence of points converging to  $x$  and let  $g_n$  be such that  $g_n(x_n) = x_n$ . Let  $z$  be any point of  $E$ . Then any limit point of the set  $g_n(z)$  is inside or on any orbit of  $G_x$  which includes  $z$  and every  $x_n$ .*

Our task now is the proof of the following theorem.

**THEOREM 6.1.** *Let  $x_n$  approach  $x$  and let  $g'_n$  be elements of  $G$  such that  $g'_n(x_n)$  approaches  $y$ . Then there is a subsequence  $g''_n$  of the  $g'_n$  and an element  $g^*$  of  $G$  such that  $g''_n$  approaches  $g^*$  (in the sense of pointwise convergence).*

The proof of this theorem is rather long and is based on a number of preliminary lemmas to which we now turn. Lemma 3.1 will also be useful here. As usual we use the letter  $g$  for an element of  $G$  and we use the letter  $h$  for a homeomorphism of  $E$  which is not known to be an element of  $G$ . Convergence of homeomorphisms, as the statement of the theorem implies, is always taken in this section in the sense of pointwise convergence, that is,  $h_n$  converges to  $h$  provided that, for each  $x$ ,  $h_n(x)$  converges to  $h(x)$ .

**LEMMA 6.12.** *Let  $g_n$  be a sequence of elements of  $G$  converging to a homeomorphism  $h$  of  $E$ . Let  $F$  be an arbitrary compact subset of  $E$  and let a positive number  $\epsilon$  be given. Then there exists an integer  $N$  such that if  $y$  and  $z$  are any two points of  $F$  for which  $d(y, z) < 1/N$ , and  $n > N$  then*

$$d[g_n(y), h(z)] < \epsilon.$$

The proof of this lemma which is quite similar to various preceding proofs will be omitted.

In the hypotheses of the following lemmas it will frequently occur as it did in the preceding lemma that there is a sequence  $g_n$  of elements of  $G$  converging to a homeomorphism  $h$  of  $E$ . From now on we shall express this fact in abbreviated form by writing  $g_n \rightarrow h$ .

**LEMMA 6.13.** *Let  $g_n \rightarrow h$  and let  $i$  be a positive integer. Then it is true that  $g_n^i \rightarrow h^i$ .*

The proof will be made by induction. It is true when  $i=1$  and we now assume that it is true for  $i-1$ .

Let  $x$  be any point of  $E$  and let  $F$  be the set made up of the points  $x$ ,  $h^{i-1}(x)$ , and  $g_n^{i-1}(x)$ , ( $n=1, 2, \dots$ ). By the hypothesis of the induction  $F$  is compact. Let  $\epsilon$  be any positive number. By the preceding lemma there is an integer  $N$  such that if  $y$  and  $z$  are in  $F$  and  $d(y, z) < 1/N$  then for  $n > N$

$$d[g_n(y), h(z)] < \epsilon.$$

On the other hand for sufficiently large  $n$ , say  $n$  greater than  $N'$ ,

$$d[g_n^{i-1}(x), h^{i-1}(x)] < 1/N.$$

Hence if  $n$  is larger than  $(N, N')$  we obtain (letting  $y$  be  $g_n^{i-1}(x)$  and  $z$  be  $h^{i-1}(x)$ )

$$d[g_n^i(x), h^i(x)] < \epsilon.$$

LEMMA 6.14. *Let  $g_n \rightarrow h$  and let  $g$  be an arbitrary element of  $G$ . Then  $gg_n \rightarrow gh$ .*

For an arbitrary  $x$  we know that  $g_n(x) \rightarrow h(x)$ . It is then an immediate consequence of the definition of homeomorphism that  $gg_n(x) \rightarrow gh(x)$ .

LEMMA 6.141. *Let  $g_n \rightarrow h$  and assume that there are elements  $g$  and  $g'$  of  $G$  such that  $gg_n \rightarrow g'$ . Then  $g' = gh$  and  $h$  is in  $G$ .*

By the preceding lemma  $gg_n \rightarrow gh$  and hence  $g' = gh$ .

LEMMA 6.15. *Let  $K$  be a simple closed curve and let  $T$  be a nonidentical sense preserving homeomorphism with a fixed point. Then there exists a pair of points  $x$  and  $y$  of  $K$  such that  $T^i(x) = x$  and  $T^i(y) \rightarrow x$  as  $i \rightarrow \infty$ .*

Choose any moving point  $y$ . Then  $T^i(y)$  must approach monotonically a point  $x$  which is fixed.

LEMMA 6.151. *Let  $g_n \rightarrow h$  and suppose that, for a definite point  $q$  of  $E$ ,  $h(q) = q$ . Then  $h$  must leave invariant every orbit of  $G_q$ , that is, for all  $x$ ,*

$$hG_q(x) = G_q(x).$$

Let  $x$  be any point. Since  $g_n \rightarrow h$  we know that  $g_n(x) \rightarrow h(x)$  and  $g_n(q) \rightarrow h(q) = q$ . Hence there is an element of  $G$  taking  $q$  to  $q$  and  $x$  to  $h(x)$ . This element is in  $G_q$  and we have thus shown that  $hG_q(x)$  is in  $G_q(x)$ . The equality must hold because of the nature of  $G_q(x)$  in the two cases.

LEMMA 6.152. *Let  $x_n$  be a sequence of points converging to a point  $x$  of  $E$ . Let  $g_n$  of  $G$  be such that  $g_n(x_n) \rightarrow x$ . Then there exists a homeomorphism  $h$  of  $E$  such that  $g_n' \rightarrow h$  for a subsequence  $g_n'$  of  $g_n$ .*

The proof here is similar to that of Lemma 4, but it depends also on Lemma 6.

LEMMA 6.16. *Let  $x_n \rightarrow x$  and let  $g_n'$  be such that  $g_n'(x_n) = y_n \rightarrow y$ . Let  $g$  (in  $G$ ) be such that  $g(y) = x$ . Then  $g_n = gg_n'$  has a subsequence  $g_n''$  which converges to a homeomorphism  $h$  and  $h(x) = x$ .*

Since  $y_n \rightarrow y$ ,  $g(y_n) \rightarrow g(y) = x$ . Then  $g_n(x_n) = gg_n'(x_n) = g(y_n) \rightarrow x$ . By the preceding lemma there is a subsequence  $g_n''$  and a homeomorphism  $h$  such that  $g_n'' \rightarrow h$ . By Lemma 3.1,  $g_n''(x_n) \rightarrow h(x)$  and hence  $h(x) = x$ .

We are now ready to prove Theorem 6.1 for the plane and we restate it here for this case.

THEOREM 6.1'. *Let  $x_n \rightarrow x$  and let  $g_n'$  be elements of  $G$  such that  $g_n'(x_n) \rightarrow y$ . Then there is a subsequence  $g_n'''$  of the  $g_n'$  and an element  $g^*$  of  $G$  such that  $g_n''' \rightarrow g^*$ . (For  $E_2$ .)*

Let  $g_n''$  and  $h$  be as in Lemma 6.16. The elements  $g_n''$  are sense preserving

and consequently  $h$  is sense preserving. For all  $z$  distinct from  $x$ ,  $G_x(z)$  is a simple closed curve. Since  $hG_x(z)$  is in  $G_x(z)$  and since  $h$  is a homeomorphism,  $hG_x(z) = G_x(z)$  and  $h$  is sense preserving on  $G_x(z)$  (see Lemma 6.151).

For convenience let  $K = G_x(z)$  for an arbitrary point  $z$  distinct from  $x$ , and let  $z' = h(z)$ . Since  $G_x$  is transitive on  $K$ , there is a  $g'$  in  $G_x$  such that  $g'(z') = z$ , that is,  $g'h(z) = z$ . By a previous lemma  $g'g_n'' \rightarrow g'h$ . Since  $g'$  is in  $G_x$ ,

$$g'h(x) = g'(x) = x.$$

It follows that  $h' = g'h$  preserves the orbits of  $G_x$  and in particular it follows that  $h'$  is a sense preserving homeomorphism of  $K$ . Moreover  $h'(z) = z$ .

We wish to prove that  $h'$  is the identity on  $E$  and we begin by proving that  $h'$  is the identity on  $K$ .

If  $h'$  is not the identity, then there is a pair of distinct points  $x^*$  and  $y^*$  such that  $h'^i(x^*) = x^*$  and  $h'^i(y^*)$  approaches  $x^*$  as  $i$  approaches infinity. For a fixed  $i$  we know that  $g_n^{*i} \rightarrow h'^i$  where  $g_n^* = g'g_n'' \rightarrow g'h = h'$ . In view of these two facts we see that  $g_n^{*i}$  for some  $n$  and  $i$  takes  $x^*$  and  $y^*$  into a specified neighborhood of  $x^*$  which violates the two-point condition. Hence  $h'$  is the identity on  $K$ .

We shall show next that the set of points of  $E - x$  on which  $h'$  is the identity contains only inner points. Let  $z'$  be any point in  $E - x$  which is fixed under  $h'$ . Since  $h'(z') = z'$ ,  $h'$  must leave invariant the orbits  $G_{z'}(z^*)$  for any  $z^*$  in  $E$ . Let  $K' = G_{z'}(z')$ . By the argument above we see that  $h'$  leaves every point of  $K'$  fixed. Let  $U$  be a neighborhood of  $z'$  so small that points of  $K'$  lie outside of  $U$  and let  $V$  be a neighborhood of  $z'$  which is invariant under  $G_{z'}$  and is contained in  $U$ . For any point  $z^*$  of  $V$  the simple closed curve  $K^* = G_{z'}(z^*)$  must contain a point of  $K'$ . This is a fixed point of  $h'$  and the argument of the preceding paragraph shows that every point of  $K^*$  is fixed under  $h'$ . Then it follows that all of  $V$  is fixed under  $h'$ .

Hence the set of points fixed under  $h'$  is both open and closed in  $E - x$ . This set of fixed points therefore includes  $E - x$  and since it is closed it must include  $E$ . Hence  $h'$  is the identity.

Now  $h' = g'h$  so that  $h = g'^{-1}$  and  $h$  is an element of  $G_x$ . This completes the proof of the theorem for the case of the plane. To see this assume for convenience that  $g_n''$  and  $gg_n'$  coincide. Then  $gg_n' \rightarrow h = g'^{-1}$  and  $g_n' \rightarrow g^{-1}g'^{-1}$ .

We turn next to the proof of the theorem for  $E_3$ .

**THEOREM 6.1''.** *Let  $x_n \rightarrow x$  and let  $g_n'$  be elements of  $G$  such that  $g_n'(x_n) \rightarrow y$ . Then there is a subsequence  $g_n''$  of  $g_n'$  and an element  $g^*$  of  $G$  such that  $g_n''' \rightarrow g^*$ . (For  $E_3$ .)*

Let  $g_n''$  and  $h$  be as in the preceding proof, that is, as in Lemma 6.16. The transformations  $g_n''$ , and therefore  $h$  also, are sense preserving in  $E$ . The point  $x$  is fixed under  $h$  and  $h$  preserves orbits  $G_x(z)$ , these orbits being two-spheres. Let  $z$  be a definite point of  $E$  distinct from  $x$  and let  $S = G_x(z)$ . Since



$h$ , being sense preserving, has at least one fixed point on  $S$  we may assume that  $z$  is such a fixed point. Then  $h$  will preserve orbits  $G_z(y)$  for all  $y$  of  $E$ .

We are going to look for some simple closed curve  $K$ , on  $S$ , invariant under  $G_z$  and  $G_z$  and also invariant under  $h$ . Let  $S'$  denote a sphere orbit under  $G_z$  such that some points of  $S$  are outside  $S'$ . Let  $F$  denote the closed intersection of  $S$  and  $S'$  and let  $D$  denote the component of  $S - F$  which contains  $z$ .

Now  $S$  and  $S'$  are invariant under  $h$  and their intersection  $F$  must also be invariant under  $h$ . Hence  $S - F$  is invariant and  $D$  must be invariant under  $h$  since it is a component with a fixed point. Let  $K$  be the boundary of  $D$  on  $S$ . It is this set which will be shown to be a simple closed curve. Observe at the moment that  $K$  is invariant under  $h$  and is a subset of  $F$ .

The group  $G_{zz}$  leaving  $x$  and  $z$  fixed is a circle group and  $F$  is invariant under  $G_{zz}$ . Let  $y$  denote any point of  $K$  above. Then  $K = G_{zz}(y)$ . Hence  $K$  is a simple closed curve invariant under  $h$ . Furthermore  $h$  in its action on  $K$  must be sense preserving for otherwise it would have to interchange the two components of  $S - K$  which it cannot do since we know that  $D$  contains a fixed point.

Since  $G_{zz}$  is transitive on  $K$  there is an element  $g'h = h'$  which has all the properties of  $h$  which we need to use and in addition has a fixed point on  $K$ .

We wish to show that  $h'$  is the identity on  $S$  ultimately enlarging the set of fixed points to include all of  $E$ . By a familiar argument  $h'$  must leave all of  $K$  fixed. The following lemma will be useful to us as we proceed.

**LEMMA 6.17.** *If  $h'$  leaves fixed a point  $p$  and a continuum on an orbit  $S^* = G_p(q)$ , then  $h'$  leaves all of  $S^*$  fixed.*

Let  $F$  be the set of fixed points of  $h'$  on  $S^*$ , and let  $M = S^* - F$ . Let  $O$  be a component of  $M$ , and let  $B$  be the boundary of  $O$ . Now from the hypothesis  $B$  cannot be zero dimensional, for if it were  $B$  would be all of  $F$  and so  $F$  would not contain a continuum. Hence  $B$  is one dimensional, and it must contain a continuum  $C$  which contains a point  $b$  which is accessible from  $O$ . Let  $U$  be a neighborhood of  $b$  (in  $S^*$ ) so small that any simple closed curve in  $U$  surrounding  $b$  must meet  $C$ . Let  $G_b(y)$  be an orbit so small that its intersection with  $S^*$  is in  $U$ . Then by an argument given above there is a simple closed curve in the intersection of  $G_b(y)$  and  $S^*$  which is invariant under  $h'$ . Furthermore  $h'$  preserves sense on this curve. Since this curve surrounds  $b$  it must meet  $C$  and consequently  $h'$  must leave all the curve fixed. As this curve was arbitrarily small we see that  $b$  cannot be accessible from  $O$ . This is a contradiction which proves the lemma.

We can of course conclude now that  $h'$  leaves all of  $S$  fixed. Furthermore any small "sphere" with center on  $S$  will also be left fixed by  $h'$  since it will intersect  $S$  in a one dimensional set. The set of "spheres" about  $x$  forms a ray (the end point of the ray being of course a point orbit). The above considera-

tions show that the "spheres" of this ray left entirely fixed by  $h'$  form a set which is both open and closed. Therefore  $h'$  leaves every point of  $E$  fixed and is the identity.

Now as in the plane case  $h = g'h$  and  $h = g'^{-1}$  so that  $h$  is in  $G$  and indeed in  $G_{xx}$ . This completes the proof as in the former case.

12.1. LEMMA 6.2. *If  $g_n \rightarrow g$ , then  $g_n^{-1} \rightarrow g^{-1}$ .*

Let  $x$  be any point of  $E$ . By hypothesis  $g_n(x) \rightarrow g(x)$ . Also  $g_n g_n^{-1}(x) \rightarrow g g^{-1}(x) = x$ . If we knew that  $g g_n^{-1}(x) \rightarrow x$  we could conclude that  $g_n^{-1}(x) \rightarrow g^{-1}(x)$ . The proof of the lemma is therefore reduced to the proof of the following special case. *If  $g_n$  is in  $G$  and  $g_n(x) \rightarrow x$ , then  $g_n^{-1}(x) \rightarrow x$ .*

In order to prove this proposition let  $S$  be any sphere about  $x$ . We shall show that almost all the points  $g_n^{-1}(x)$  are inside  $S$ . If this is not true we arrive at a contradiction as follows. For each  $g_n^{-1}(x)$  not inside  $S$  choose a short arc joining  $x$  to  $g_n(x)$ , call it  $xg_n(x)$ . Then  $g_n^{-1}(xg_n(x))$  will be an arc joining  $g_n^{-1}(x)$  to  $x$ . It will therefore contain a point  $y_n$  on  $S$ . Now  $g_n$  takes the pair  $(y_n, x)$  to a pair  $(g_n(y_n), g_n(x))$  both elements of this pair being near  $x$ . The points  $y_n$  will have a limit  $y$  on  $S$  and there will have to be a  $g$  in  $G$  taking both  $y$  and  $x$  to  $x$ . This contradiction establishes the lemma.

LEMMA 6.3. *If  $g_n \rightarrow g$  and  $g'_n \rightarrow g'$ , then  $g_n g'_n \rightarrow gg'$ .*

Let  $x$  be any point of  $E$ . Then  $g'_n(x) \rightarrow g'(x)$ . Hence  $g_n[g'_n(x)] \rightarrow gg'(x)$  which we wanted to prove.

THEOREM 6.2. *Let  $(a_1, \dots, a_k)$  and  $(A_1, \dots, A_k)$  be any two sets of  $k$  points. Let  $(a_1^n, \dots, a_k^n)$  approach  $(a_1, \dots, a_k)$  and let  $(A_1^n, \dots, A_k^n)$  approach  $(A_1, \dots, A_k)$ . If there is for each  $n$  an element  $g_n$  in  $G$  such that*

$$g_n(a_i^n) = A_i^n, \quad i = 1, \dots, k,$$

*then there exists an element  $g$  in  $G$  such that*

$$g(a_i) = A_i, \quad i = 1, \dots, k.$$

Let  $g'_n$  be an element of  $G$  such that  $g'_n(a_i) = A_i^n$ . Then  $g_n g'_n(a_i) = A_i^n \rightarrow A_i$ . Letting  $i=1$  we see that  $g_n g'_n(a_1) \rightarrow A_1$ . Since  $a_1$  certainly approaches  $a_1$ , Theorem 6.1 tells us that the sequence  $g_n g'_n$  contains a subsequence converging to an element  $g^*$  in  $G$ . We assume that our original sequence above is taken as this convergent subsequence. An application of Theorem 6.1 to the sequence  $g'_n$  (remembering that  $g'_n(a_1) = A_1^n \rightarrow A_1$ ) shows that for a subsequence of  $n$ 's (which we now take for the whole sequence)  $g'_n \rightarrow g'$  where  $g'$  is in  $G$ . We have now arrived at the following situation:

$$g_n g'_n \rightarrow g^*, \quad g'_n \rightarrow g'.$$

By the preceding results  $g_n'^{-1} g_n^{-1} \rightarrow g^{*-1}$ . Hence



$$g'_n(g_n'^{-1}g_n^{-1}) = g_n^{-1} \rightarrow g'g'^{-1}, \quad g_n \rightarrow g^*g'^{-1}.$$

Now for any  $i$  ( $1 \leq i \leq k$ )

$$A_i^n = g_n(a_i) \rightarrow g^*g'^{-1}(a_i).$$

But  $A_i^n \rightarrow A_i$  and hence

$$g^*g'^{-1}(a_i) = A_i.$$

The element  $g^*g'^{-1}$  therefore has the desired properties.

This theorem applied to the plane case (and with  $k=3$ ) yields Hilbert's Axiom III. It is worth pointing out that we have actually proved a great deal more than Hilbert's axioms; we have also obtained many of the results of his paper. But as we are, at the moment, only interested in establishing that our weaker axioms suffice for the plane, we leave the plane case and turn our entire attention to the three dimensional case. From now on it is to be understood that we are dealing with this latter case.

13. By a *G-straight*, or more simply a *straight* or a *line* we shall mean a topological line which is the set of points left fixed by a circular subgroup of some  $G_x$ . Through every point of space there is clearly a large family of straights. Now let  $x$  and  $y$  be any two distinct points of the space. In the group  $G_x$  there is one and only one circular subgroup leaving  $y$  fixed. Let this group be called  $K_{xy}$ . It clearly leaves fixed a topological line, call it  $L_{xy}$ . We have therefore seen that there is at least one *G-straight* through every two distinct points of space. We know that  $L_{xy}$  is the set of all fixed points of  $K_{xy}$ . We have, therefore, the following theorem.

**THEOREM 7.** *Through each pair of distinct points of  $E$  there passes one and only one  $G$ -straight.*

13.1. It is worth noting that if an element  $g$  in  $G$  leaves fixed three points  $x, y, z$  not all on the same straight then  $g$  leaves all of  $E$  fixed. For since  $g$  leaves  $x$  fixed it is in  $G_x$ . Since it leaves  $y$  fixed it is in the circular subgroup  $K_{xy}$  of  $G_x$ ; and since it leaves fixed  $z$ , a point not on the "axis" of  $K_{xy}$ , it must leave all of  $E$  fixed.

The symbol  $xy$  will be used to denote the closed portion of the line  $L_{xy}$  which is contained between the points  $x$  and  $y$ . This set of points will be called an *interval* or the *interval*  $xy$ , or a *segment*.

Let  $L_{xy}$  be a straight, left fixed by the circular group  $K_{xy}$ , and let  $g$  be any element of  $G$ . Then the set of points  $g(L_{xy})$  is the set left fixed by  $gK_{xy}g^{-1}$ . In other words the image of a straight under any element of  $G$  is also a straight. It follows that the image of a segment is a segment.

Two configurations of the space  $E$  are said to be *congruent* if one of them is carried into the other by some element of  $G$ .

Any two straights are congruent and in fact any two marked straights are

congruent. By a *marked straight* we mean a straight with some one of its points particularly "marked." It is furthermore true that any marked straight can be taken to any other so that a given direction on the one goes to a given direction on the other. These facts follow from the transitivity of  $G$  and the nature of the rotation group.

14. As we have said, a sphere is defined to be any two dimensional orbit of any group  $G_x$ . The point  $x$  is called the center of the sphere.

**THEOREM 8.** *A straight and a sphere can intersect in two points at most.*

Let  $x$  be the center of  $S$  the sphere and let  $L$  be the straight. Suppose  $a$  and  $b$  are two points of intersection of  $L$  and  $S$ . There exists an element of  $G_x$  which interchanges  $a$  and  $b$ . This element must carry  $L$  into itself since  $L$  is determined by any two of its points. Then there exists a non-trivial subgroup of  $G_x$  which leaves  $L$  invariant. This subgroup which will be denoted by  $Q$  is compact. Such a compact group acting on a line can contain only two distinct transformations, the identity and a reflection. Hence under this group the orbit of  $a$  consists of the two points  $a$  and  $b$ . It follows that there can be no other point  $c$  of  $L$  on  $S$ , for otherwise there would be an element of  $Q$  interchanging  $a$  and  $c$  and  $a$  would have at least three points in its orbit. This completes the proof.

**THEOREM 9.** *If  $p$  and  $q$  are inside a sphere  $S$ , then the segment  $pq$  is inside  $S$ .*

The straight  $L_{pq}$  is not compact in either direction. In going along  $L_{pq}$  from  $p$  to  $q$  and on we must meet  $S$  in some point. Similarly we must meet  $S$  in going from  $q$  to  $p$  and on out. The straight  $L_{pq}$  meets  $S$ , then, in two points neither of which is in the interval  $pq$ . Therefore no point of the interval can be outside  $S$  for this would imply that some point of the interval was on  $S$  and this would mean that  $L_{pq}$  had at least three distinct points on  $S$ .

15. **THEOREM 10.** *If  $x_n \rightarrow x$  and  $y_n \rightarrow y$  then  $x_n y_n \rightarrow xy$ .*

Let  $z_n$ ,  $n = 1, 2, 3, \dots$ , be a point of  $x_n y_n$ . Any sphere surrounding  $xy$  surrounds almost all the  $z_n$ , so that for a proper subsequence of the  $n$  and a suitable point  $z$ ,  $z_n \rightarrow z$ . We have to prove that  $z$  is on  $xy$ . There exist elements  $g_n$  such that  $g_n x_n = x$ , and  $g_n y_n$  is on  $L_{xy}$  on the same side of  $x$  as the point  $y$ .

Now we know that there is a subsequence of the  $g_n$  and an element  $g$  such that  $g_n \rightarrow g$ . Further  $g x = x$ ,  $g y_n \rightarrow g y$ , and  $g z_n \rightarrow g z$ . Now  $g y_n$  is on  $L_{xy}$  so that  $g y$  must also lie on  $L_{xy}$ . On the other hand  $g y$  belongs to  $G_x(y)$ . Then it is clear that  $g y = y$ . Therefore  $g$  belongs to  $K_{xy}$ ,  $g^{-1}$  belongs to  $K_{xy}$  and  $z = g^{-1}(g z)$  is a point of  $L_{xy}$  since  $g z$  is a point of  $L_{xy}$ . It is now a trivial matter that  $z$  is on  $xy$  and that every point of  $xy$  is a limit point of some sequence  $z_n$ .

It should be remarked that under the same hypothesis the line  $L_{x_n y_n}$  converges to the line  $L_{xy}$ . By this we mean that every sequence  $z_n$  of points from

$L_{x_n y_n}$  either has no limit point or every limit point which it has is on  $L_{xy}$ . Furthermore every point of  $L_{xy}$  is a limit point of such a sequence.

Let us now take a point  $z_n$ ,  $n = 1, 2, 3, \dots$ , on  $L_{x_n y_n}$  and assume that the sequence  $z_n$  converges to a point  $z$ . The intervals  $x_n z_n$  then converge to  $xz$  and  $y_n z_n$  converge to  $yz$ . It may be assumed that the points  $z_n$  are outside the interval  $x_n y_n$  say in the order  $x_n y_n z_n$ . The segment  $xz$  then contains  $xy$  and  $yz$ , so that  $xz$  is clearly part of the line  $L_{xy}$ .

The argument that every point of  $L_{xy}$  is such a limit is not difficult.

16. THEOREM 11. *Let  $x$  and  $y$  be any two points of the sphere  $S$ . Then the interval  $xy$ , except for  $x$  and  $y$ , is inside  $S$ .*

Let  $x_n$  and  $y_n$  be sequences of points inside  $S$  converging respectively to  $x$  and  $y$ . By a preceding theorem the intervals  $x_n y_n$  must be contained in the interior of  $S$ . The limit of these intervals will then be inside or on  $S$ . This limit is  $xy$ . This shows that  $xy$  is entirely contained in  $S$  and its interior. But  $L_{xy}$  can have no point besides  $x$  and  $y$  on  $S$ . The conclusion therefore follows.

17. A point  $y$  is said to be the midpoint of the segment or interval  $xz$  if it is on  $xz$  and if there is an element of  $G$  leaving  $y$  fixed and interchanging  $x$  and  $z$ . The point  $y$  is then the center of a sphere containing  $x$  and  $z$  as antipodal points.

THEOREM 12. *Every segment has a unique midpoint.*

We begin by showing the existence of the midpoint. Let  $y$  be a variable point of the segment  $xz$ . There is in  $G_y$  an element of order two which moves  $x$  to a unique point of  $L_{xz}$ . Let this unique point be denoted by  $f(y)$ . Now  $f(x) = x$ ; and  $f(z)$  lies on  $L_{xy}$  and has the order  $xzf(z)$ . Therefore if  $f(y)$  is continuous it will assume for some  $y$  the value  $z$ . We have only to show therefore that  $f(y)$  is continuous.

Let  $y_n$  approach  $y_0$ . Lemma 6 shows that  $f(y_n)$  is a bounded set, and we may assume that  $f(y_n)$  approaches a point  $w$  of the line  $L_{xz}$ . We wish to show that  $w = f(y_0)$ . Assume that this is not true and that  $w$  is distinct from  $f(y_0)$ . The pair  $(x, y_n)$  is carried by an element of the group to the pair  $[f(y_n), y_n]$ . There must be an element  $g$  of the group taking the pair  $(x, y_0)$  to the pair  $(w, y_0)$ . There is also an element  $g'$  taking the pair  $(x, y_0)$  to the pair  $(f(y_0), y_0)$ . That is

$$g(y_0) = g'(y_0) = y_0$$

and

$$g(x) = w, \quad g'(x) = f(y_0).$$

Hence  $g'g^{-1}$  leaves  $y_0$  fixed and takes  $w$  to  $f(y_0)$ . This is impossible because  $g'g^{-1}$  is in the compact group  $G_{y_0}$  and the points  $w$  and  $f(y_0)$  are both on the

same side of  $y_0$  on the line  $L_{xz}$ . Hence  $f(y)$  is continuous and the midpoint of  $xz$  exists.

Assume now that there are two midpoints  $y$  and  $y'$  of the segment  $xz$ . Each of them gives rise to a reflection interchanging  $x$  and  $z$  and leaving itself fixed. The product of these two reflections of  $L_{xz}$  is a transformation of  $L_{xz}$  leaving  $x$  and  $z$  fixed and moving other points on the line. This is impossible.

18. The next geometric concept to be defined will be the *projection* of a point  $z$  on a line  $L$ . If  $z$  is on  $L$  this projection is defined to be  $z$  itself. If  $z$  is not on  $L$  let  $g$  be an element of order two in  $K_L$  which is the circular group leaving every point of  $L$  fixed. Let  $z' = g(z)$ ; evidently  $g(z') = z$ . The line  $L_{zz'}$  and the segment  $zz'$  are invariant under  $g$ , so that  $g$  is a reflection of the line and segment with a fixed point  $p$ . Since this point  $p$  is fixed under an element of  $K$  not the identity, it is fixed under all of  $K$  and is on  $L$ . The point  $p$  is now defined to be the projection of  $z$  on  $L$ .

**THEOREM 13.** *The projection of  $z$  on  $L$  is a continuous function of  $z$ .*

Suppose  $z_n$  converges to  $z$ . Then  $g(z_n) = z'_n$  converges to a point  $z'$ ,  $g$  being the element of order two in the group leaving  $L$  fixed. The segments  $z_n z'_n$  converge to the segment  $zz'$ . The points  $p_n$  have some limit point on  $zz'$ . But a limit point of the  $p_n$ 's must be fixed under  $g$  and can only be the point  $p$ .

19. The space  $E$  is given to us as a metric space. Our purpose now is to introduce a new metric equivalent to the old which will be invariant under  $G$ .

Let  $L$  be any straight in  $E$  and let  $G^*$  be the set of elements of  $G$  which transform  $L$  into itself while preserving direction. Any element of  $G^*$  which leaves a point of  $L$  fixed leaves every point of  $L$  fixed. Let  $g_n$  be a sequence of elements of  $G^*$  such that for some point  $a$  in  $L$  the sequence  $g_n(a)$  approaches  $a$ . Then for any  $b$  in  $L$  the sequence  $g_n(b)$  approaches  $b$ . It follows from the fact mentioned here that if two homeomorphisms of  $G^*$  act approximately the same way on a single point of space then they act approximately the same way over any bounded part of the line.

It will be useful to note also that if  $ab$  is an interval of  $L$  and  $g$  takes  $L$  into itself with direction reversed and if  $g(b) = a$  it follows that  $g(a) = b$ . With the aid of  $G^*$  we shall now see how  $L$  becomes the carrying space of a topological group.

19.1. Let  $o$  be a point of  $L$  fixed but arbitrary. Let  $a$  and  $b$  be any two points of  $L$ . Let  $f$  be an element of  $G^*$  which moves  $o$  to  $a$ , and let  $g$  be an element of  $G^*$  which moves  $o$  to  $b$ . Then by definition  $a \cdot b = fg(o)$ . This definition of multiplication on  $L$  is associative. It has an inverse, and hence all the group axioms are satisfied. If  $f(o) = a$ , the inverse of  $a$  is  $f^{-1}(o)$ . Another way to obtain  $a^{-1}$  is to define it as the position to which  $a$  goes by an element which reverses direction on  $L$  while leaving  $o$  fixed. The latter definition shows that the operator "inverse" is continuous.

Now assume that  $a_n$  approaches  $a$  and  $b_n$  approaches  $b$ . Let  $f_n(o) = a_n$ ,

$g_n(o) = b_n$ ,  $f(o) = a$ , and  $g(o) = b$ . By the remarks in §19  $f_n$  is near  $f$  for large  $n$  and any bounded portion of the line. A similar statement may be made about  $g_n$  and  $g$ . Therefore  $f_n g_n(o)$  is near  $fg(o)$ . This shows that the group multiplication  $a \cdot b$  is simultaneously continuous in  $a$  and  $b$ .

Thus with multiplication as above defined  $L$  becomes the carrier of a topological group. But it is known that such a group must be bicontinuously isomorphic to the additive group of real numbers. The line  $L$  may now be metrized with the metric of the real numbers carried by it in this way. That is, if  $x$  and  $y$  are any two points of  $L$ , then  $d^*(x, y)$ , the new distance, is defined to be the absolute value of the difference of the real numbers corresponding to  $x$  and  $y$ .

Now suppose that the segment  $xy$  is translated to  $x'y'$  by an element  $g$  of  $G^*$ . Thus:

$$g(x) = x', \quad g(y) = y'.$$

If  $g(o) = z$ , then

$$z \cdot x = x', \quad z \cdot y = y'.$$

Therefore  $x \cdot y$  is translated to  $x' \cdot y'$  by an operation of the topological group defined above and therefore  $d^*(x, y) = d^*(x', y')$ .

Any element of  $G$  which reverses sense on  $L$  also preserves the new distance. This is because there is one sense reversing transformation which merely changes the ends of a given interval. This transformation leaves the length of the interval invariant. Any other sense reversing transformation is the product of this one and an element of  $G^*$ , from which our statement now follows.

19.2. We may now define the new distance for any two points  $x$  and  $y$  of space. There is some element in  $G$  which carries  $x$  and  $y$  to two points  $x'$  and  $y'$  of  $L$ . The distance  $d^*(x, y)$  is defined as equal to  $d^*(x', y')$ . By its very definition this new distance extended as it now is to all of  $E$  is invariant under  $G$ . In fact any two pairs of points are congruent if and only if they have the same distance  $d^*$ .

19.3. To show that the new distance  $d^*$  is equivalent to the old it is only necessary to show that it is a continuous function.

**THEOREM 14.** *The distance  $d^*(x, y)$  is a continuous function of  $x$  and  $y$ .*

Let  $x_n$ 's converge to  $x$  and  $y_n$ 's to  $y$ . We saw in the proof of Theorem 10 that  $G$  contains an element which takes  $x_n$  to  $x$  and  $y_n$  to the line  $L_{xy}$ , near  $y$ . But on  $L$  and hence on  $L_{xy}$ ,  $d^*$  is a continuous function. This shows that  $d^*$  of the transformed pair, which is the same as  $d^*$  for the original pair  $(x_n, y_n)$ , is near  $d^*(x, y)$ .

19.4. **THEOREM 15.** *The distance  $d^*(x, y)$  satisfies the triangle axiom.*



Let  $A$ ,  $B$ , and  $C$  be any three points of  $E$ . We wish to prove that

$$d^*(A, C) + d^*(C, B) \geq d^*(A, B).$$

If  $C$  is on the line  $L_{AB}$  we know this. When  $C$  is between  $A$  and  $B$  the equality holds, and this will be the only case of equality.

Suppose now that  $C$  is not on  $L_{AB}$  and let  $C'$  be the projection of  $C$  on the line. To prove the desired inequality it will suffice to show that  $d^*(A, C) > d^*(A, C')$  and  $d^*(C, B) > d^*(C', B)$ . This follows from the following lemma.

LEMMA 7. *Let  $x$ ,  $y$ , and  $z$ , not on a line, be such that  $z$  is the projection of  $y$  on  $L_{xz}$ . Then  $d^*(x, y) > d^*(x, z)$ .*

Let  $S$  be the sphere about  $x$  which goes through  $y$ , that is, the sphere with center  $x$  and radius  $d^*(x, y)$ . Let  $g$  be the element of order two in  $K_{xz}$  and let  $y' = g(y)$ . The segment  $yy'$  has  $z$  as midpoint, so that certainly  $z$  is on  $yy'$ . The point  $y'$  is on  $S$  because  $xy$  and  $xy'$  are congruent under  $g$ . If now  $d^*(x, z) = d^*(x, y)$ , there would have to be a point  $z'$  on the segment  $xz$  such that  $d^*(x, z') = d^*(x, y)$ . Since  $y$  and  $y'$  are on the sphere,  $z$  is inside it. Of course  $x$  is inside the sphere and hence  $z'$  is also inside it, contrary to the preceding equality.

It should be noted that all points of  $yy'$  except  $y$  and  $y'$  are inside  $S$  so that for any point  $w$  of  $yy'$  distinct from  $y$  and  $y'$ ,  $d^*(x, w) < d^*(x, y)$ .

It is clear that we have established the metric characterization of a straight line: three points are on a straight line if and only if their distances in proper and unique order satisfy the triangle equality. From this point on, the distance  $d^*$ , now called  $d$ , will be the only distance used.

20. Let  $L$  and  $L'$  be two lines having in common the single point  $p$  and such that there exists an element of order two in  $K_L$  under which the line  $L'$  is invariant. In this case  $L'$  is said to be *orthogonal* or *perpendicular* to  $L$ . If  $x$  is a point on  $L'$ , its projection on  $L$  is the point  $p$ .

Consider the rotation group  $G_p$  and the coordinate system that goes with it which makes  $G_p$  the three-space rotation group. Since there is an element of order two, that is, a half-rotation in  $G_p$  which leaves  $L$  fixed and  $L'$  invariant, there is also a half-rotation in  $G_p$  which leaves  $L'$  fixed and  $L$  invariant. In other words  $L$  is also orthogonal to  $L'$  and the relation of orthogonality is symmetric. It can be seen also that orthogonality is a group invariant.

20.1. The transitivity of  $G$  and the nature of the rotation group enables us to state the following two theorems.

THEOREM 16. *Let  $L$  and  $N$  be two lines both orthogonal to a line  $M$  at a point  $p$ . There is then an element of  $K_M$  which carries  $L$  to  $N$ .*

This makes it clear incidentally that the locus of points on lines orthogonal to  $L$  at  $p$  is a topological plane.



**THEOREM 17.** *Let  $L$  and  $L'$  be orthogonal at  $p$  and let  $M$  and  $M'$  be orthogonal at  $q$ . There is then an element  $g$  in  $G$  such that  $g(p)=q$ ,  $g(L)=M$  and  $g(L')=M'$ . Furthermore this element may be so chosen as to take any desired directions on  $L$  and  $M$  to any desired directions on  $L'$  and  $M'$ .*

20.2. Let  $L$  and  $L'$  be a pair of orthogonal lines intersecting in a point  $p$ . Let  $x$  and  $y$  be two points of  $L'$  on opposite sides of  $p$  and at equal distances from  $p$ . A half-rotation about  $L$  must then interchange  $x$  and  $y$ . If  $q$  is any point of  $L$ , the half-rotation carries the segment  $qx$  to  $qy$ , so that  $q$  is equidistant from  $x$  and  $y$ . We may express this result by saying that if two lines are orthogonal any point on one of them is equidistant from any pair of symmetrically placed points on the other.

20.3. A triangle is defined in the natural way as the system of three segments joining pairs of a set of three points. Other simple geometric concepts will sometimes be used without definition when the definition is perfectly straightforward.

As a sort of converse to the result of the preceding section the following theorem is given at this point.

**THEOREM 18.** *The altitude of an isosceles triangle bisects the base.*

Let  $q$ ,  $x$ , and  $y$  be three points such that  $qx$  and  $qy$  are congruent. Let  $p$  denote the projection of  $q$  on  $L_{xy}$ . Then  $L_{pq}$  is orthogonal to  $L_{xy}$ . Consider the half rotation which leaves  $L_{pq}$  fixed and  $L_{xy}$  invariant. This carries  $x$  and  $y$  to  $x'$  and  $y'$ , all four points being on the same line. All four segments  $qx$ ,  $qy$ ,  $qx'$ , and  $qy'$  are equal and the line  $L_{xy}$  must meet the sphere  $G_q(x)$  in four points,  $x$ ,  $y$ ,  $x'$ , and  $y'$ . At most two of these are distinct and we know that  $x$  and  $y$  are distinct. This makes it clear that  $y'$  is  $x$  and  $x'$  is  $y$ , so that  $p$  is the midpoint of the segment  $xy$ .

21. Let  $L$  denote a line and  $p$  a point on it. Let  $\pi$  denote the set of all points of space whose projection on  $L$  is the point  $p$ . Such a set, by definition, is a plane of our geometry. It is clear that any line  $M$  orthogonal to  $L$  at the point  $p$  belongs to  $\pi$ , and that every point of  $\pi$  is on one such line. Now let  $x$  and  $x'$  be a pair of points symmetrically situated about  $p$  on the line  $L$ . Then all points of  $\pi$  are equidistant from  $x$  and  $x'$ : conversely any point equidistant from these must lie on  $\pi$ . Our planes may therefore be characterized as the locus of points equidistant from some pair of points. If we consider the circle group which leaves fixed the line  $L=L_{xx'}$ , we see that every line  $M$ , as above, is generated by this group from any arbitrary one. It follows at once that our planes have the topological structure which they should.

21.1. To show that our planes are linear sets we shall borrow from Kerékjártó the notion of introducing a simple antipodal transformation  $\alpha$ , not an element of  $G$ , defined on the whole space as follows. Under  $\alpha$  the point  $p$  is fixed: a point  $q$  goes to that point  $q'$  on the line  $pq$  which is symmetrically

disposed about  $p$ . It is obvious that the straight lines through  $p$ , and only these, are invariant under  $\alpha$ .

Our metric is invariant under this transformation  $\alpha$ . To see this, following Kerékjártó, we need merely show that, for an arbitrary pair of points,  $\alpha$  coincides with an appropriate element of  $G$ . To this end, let  $q$  and  $s$  be two points, distinct from  $p$ , and let  $L'$  denote a line orthogonal to  $pq$  and  $ps$ . Let  $g$  denote a half-turn about  $L'$ . It is clear that  $q$  and  $s$  have the same images under this half-turn as under the antipodal transformation  $\alpha$ . It now follows immediately from the invariance of our metric that  $\alpha$  carries straight lines of space to straight lines, for these as we have seen are metrically characterized.

21.2. Consider now the line  $L$  and the plane  $\pi$  orthogonal to it at the point  $p$ . Perform upon space the antipodal transformation  $\alpha$  followed by a half-turn about  $L$ . It is clear that the points of  $\pi$  are fixed points under this product-transformation. But it is important to observe that they are the only fixed points: this follows from the fact that all lines through  $p$  are invariant under  $\alpha$  and only those remain invariant under the half-turn which are orthogonal to  $L$  at  $p$ . Now take two points,  $q$  and  $s$ , in  $\pi$ . The line through these must be invariant because our composite transformation carries it to a line through  $q$  and  $s$ , since these are fixed points. This transformation, moreover, preserves distances. Since it has a pair of fixed points it must leave all of the line fixed. Therefore the line through  $q$  and  $s$  must lie in  $\pi$ , the locus of fixed points. Then we have shown that with every pair of its points our plane contains the line determined by these points.

22. As we know, the plane  $\pi$  is invariant under the circle group  $K_L$ . We want to show next that associated with every point  $q$  of  $\pi$  there is a similar circle group leaving  $\pi$  invariant. Such a step will enable us to see that there is nothing special about the point  $p$  and to conclude that  $\pi$  is either a euclidean or hyperbolic plane under the subgroup  $G^*$  of  $G$  which leaves  $\pi$  invariant. This group is transitive on  $\pi$ .

22.1. Let us introduce the notion of the projection of a point  $x$  on the plane  $\pi$ . This is defined as that point  $x'$  of  $\pi$  which is nearest to  $x$ . In order to see that  $x'$  is determined, let  $z$  denote a point of  $\pi$  and let  $S$  denote a sphere with center at  $x$  and radius  $xz$ . The solid sphere intersects  $\pi$  in a compact set. For any point of  $\pi$  not in this intersection the distance to  $x$  must exceed  $xz$ . On the compact set there certainly is one "nearest" point, but conceivably more than one. Now there can be at most one such nearest point. For suppose  $x'$  and  $x''$  are two points of  $\pi$  at the same distance from  $x$ . Then the midpoint of  $x'x''$  is in  $\pi$  and is nearer to  $x$  than  $x'$  and  $x''$  are. It is clear from the continuity of distance and the uniqueness of projection that this projection operation is continuous. We shall use this continuity in the following theorem.

**THEOREM 19.** *Every point  $q$  of  $\pi$  is the projection of at least one point  $q'$  not on  $\pi$ .*

Consider a sphere about  $P$  large enough to have the point  $q$  inside of it. This sphere meets  $\pi$  in a circle, call it  $C$ . Let  $S$  denote one of the hemispheres associated with  $C$ . Let  $S^*$  denote the set of projections of  $S$ . Since  $C$  can be deformed on  $S$  to a point of  $S$ , it can be deformed on  $S^*$  to a point of  $S^*$ . During this deformation it must meet  $q$  since  $q$  is in the domain bounded by  $C$ . This means that  $q$  is a point of  $S^*$  as was to be shown.

22.2. Let  $q'$  be a point not on  $\pi$  and let  $q$ , distinct from  $p$ , be its projection on  $\pi$ . Let  $\pi'$  denote the plane of lines orthogonal to  $qq'$  at  $q$ . This plane contains all straight lines of which it contains a pair of points, by 21.2.

**THEOREM 20.** *The planes  $\pi$  and  $\pi'$  are identical.*

It will first be shown that  $\pi'$  includes  $\pi$ . Consider the line  $L_{pq}$ . This line is in both planes. The coordinate system around  $P$  shows us that there is one and only one line in  $\pi$  which goes through  $p$  and is orthogonal to  $L_{pq}$ . Let  $L'$  denote any line, distinct from this one, through  $p$  and in  $\pi$ . Let  $s$  denote the projection of  $q$  on the line  $L'$ . The point  $s$  is distinct from  $p$  by our choice of  $L'$ . Since  $s$  and  $q$  are points of  $\pi$ ,  $L_{sq}$  belongs to  $\pi$ . Since it is a line of  $\pi$  which goes through  $q$  it is also in  $\pi'$ . Therefore  $s$  and  $p$  are both points of  $\pi'$  and  $L_{sp}$  belongs to  $\pi'$ . The set of points on such lines is dense in  $\pi$ , and the closure of this set, which is  $\pi$ , must also belong to  $\pi'$ .

We will now show that  $\pi$  contains  $\pi'$ . The plane  $\pi$  has the property that each of its points is interior to a two-cell, and it must therefore be an open set in  $\pi'$ . On the other hand it is a closed subset of space and is therefore closed in  $\pi'$ . Then it must coincide with  $\pi'$ .

23. Let  $(\pi, p)$  denote a *marked plane*, that is to say a plane  $\pi$  with some one of its points  $p$  particularly specified.

**THEOREM 21.** *Any two marked planes are congruent.*

Let  $(\pi, p)$  and  $(\sigma, s)$  be the two marked planes. Let  $L$  denote a line orthogonal to  $\pi$  at  $p$ . Such a line shall be by definition a line through  $p$  orthogonal to every line of  $\pi$  through  $L$ . This orthogonal line always exists and is unique, for it is the locus of points  $p'$  which project on  $p$  when they are projected on  $\pi$ . This we will see as follows. Let  $L^*$  be the totality of points projecting on  $p$ . As we know  $L^*$  must contain at least one point  $p'$  distinct from  $p$ . Hence from previous considerations  $L^*$  must be the line  $L_{pp'}$ .

Now let  $L'$  denote a line orthogonal to  $\sigma$  at  $s$ . The marked line  $(L, p)$  may be carried to the marked line  $(L', s)$ . Since these lines completely determine  $\pi$  and  $\sigma$ ,  $(\pi, p)$  must go into  $(\sigma, s)$  by the element which takes  $(L, p)$  to  $(L', s)$ . If we carry the line  $L'$  into itself by any half turn about the point  $s$ , the plane  $\sigma$  must go into itself with orientation reversed. The marked planes are therefore congruent with a matching of any orientations that we choose on them.

24. The geometric concepts have now been analyzed sufficiently for us to

be able to see that they satisfy the axioms usually given for a geometry, with the exception of the parallel axiom. Axioms I, II, III, and V as given by Hilbert [4] for example are all satisfied.

It is clear that there are two possible geometries satisfying our axioms for the space case, the euclidean and the hyperbolic. We now sketch rapidly one method for seeing that there are not more than two.

Let  $\pi$  be any plane in  $E$ . The subgroup of  $G$  which takes  $\pi$  into itself and preserves orientation on  $\pi$  can be seen to satisfy the axioms of Hilbert's paper [5]. The plane is therefore either euclidean or hyperbolic.

Since all planes are congruent to any given plane, we see that either every plane is hyperbolic or every plane is euclidean. We wish to show that the geometry induced by  $G$  is either euclidean or hyperbolic according to the character of the planes.

Let  $(E, G)$  and  $(E', G')$  be two systems satisfying all the axioms for the space case and assume that in the two systems planes are of the same character.

Let  $(\pi, p)$  be a marked plane in  $E$  and let  $(\pi', p')$  be a marked plane in  $E'$ . Let  $H$  denote a congruence correspondence between these two planes. Let  $L$  and  $L'$  be the unique lines of  $E$  and  $E'$  which are orthogonal to  $\pi$  and  $\pi'$  at  $p$  and  $p'$ . The unit of length gives us a unique correspondence between  $L$  and  $L'$ , the only choice, and it is an arbitrary one, being which half of  $L$  is mapped on which of  $L'$ . Assume that this choice has been made so that we have really chosen an upper and a lower half for  $E$  and also for  $E'$ .

We can now choose coordinates in  $E$  and  $E'$  and extend  $H$  by letting points with the same coordinates correspond. The correspondence  $H$  as thus extended is isometric. It preserves segments, orthogonality, lines, planes, and in fact all geometric concepts. The function  $H$  also associates with every  $g$  in  $G$  an element  $g'$  in  $G'$  and we are therefore led to the conclusion that  $(E, G)$  and  $(E', G')$  are equivalent provided the planar character of the two systems is the same.

#### APPENDIX

For the space case an alternative set of axioms might be chosen as follows. Let  $(E_s, G)$  be a system consisting of a set  $G$  of sense preserving homeomorphisms of  $E_s$ . Let the following axioms be satisfied.

2.1''. The same as 2.1'.

2.2''. There is a point  $p$  such that  $G_p$  satisfies the conditions (a), (b), and (c):

(a)  $G_p$  is a proper subgroup of  $G$ .

(b) For each  $x$  distinct from  $p$ ,  $G_p(x)$  contains at least three points.

(c) For a sequence of points  $p_n$  approaching  $p$ ,  $G_p(p_n)$  is at least one dimensional.

2.3''. The same as 2.3'.

We will show that these axioms imply 2.2' from which it follows that they suffice for the foundation of space geometry.

Quite as in the paper we can arrive at the situation of §5. We have then an open set  $R_1$  with a closure  $\bar{R}_1$ , and  $\bar{H}_1$  is a compact effective transformation group of  $\bar{R}_1$ . The orbits of  $\bar{H}_1$  are the same as those of  $G_p$ .

Let  $H_1^*$  be the component of the identity of  $\bar{H}_1$ . The set  $H_1^*(x)$  has the same dimension as  $\bar{H}_1(x)$  and  $\bar{H}_1$  must have orbits of dimension at least one in  $\bar{R}_1$ .

Assume now that  $H_1^*$  is one dimensional. Then  $H_1^*$  is the circle group. As  $H_1^*$  leaves fixed an "axis" of points of  $\bar{R}_1$ , not every  $x$  distinct from  $p$  has 3 points in its orbit under  $H_1^*$ . Therefore  $H_1^*$  does not exhaust  $\bar{H}_1$ . But  $H_1^*$  divides  $\bar{R}_1$  locally (near  $p$ ) into a decomposition space which is essentially a half plane.  $\bar{H}_1 - H_1^*$  must act on this half plane and the only compact group which can act on a half plane is a group of order two which reflects its edge. Hence in any case points on the "axis" of  $H_1^*$  will have orbits of at most two points under  $\bar{H}_1$ . This shows that  $\bar{H}_1$  cannot be one dimensional. But since it is now seen to be of dimension greater than one it must also have orbits of dimension greater than one by arguments in our earlier papers. This concludes the reduction of the present system of axioms to those of this paper.

It should be remarked here that in the presence of condition (b) above it is altogether likely that condition (c) can be relaxed perhaps merely to assert, with Hilbert, that  $G_p(p_n)$  is infinite. This appears to have all of the difficulty which attends the problem of showing, if it is true, that a zero dimensional topological transformation group of three-space is necessarily finite.

We might mention in conclusion that it seems to us of some interest to determine the three-space geometries through appropriate reflection groups, along the lines on which this was done for the plane by Cairns. While it is clear that suitable conditions on "reflections" of three-space could be made to yield the axioms of this paper, the characterization of the fixed points of reflection of three-space by P. Smith might lead to an interesting approach.

Added in proof: Kerékjártó has informed us that he has published a further paper on this same subject in the Proceedings of the Hungarian Academy of Sciences (1928) (in Hungarian). He is about to publish another paper on this subject in the Acta Mathematica.

#### BIBLIOGRAPHY

1. Alexandroff and Hopf, *Topologie I*, Berlin, 1935.
2. Cartan, *La Théorie des Groupes Finis et Continus et l'Analysis Situs*, Mémorial des Sciences Mathématiques, vol. 42.
3. Cairns, *An axiomatic basis for plane geometry*, these Transactions, vol. 35 (1933), pp. 234-244.
4. Hilbert, *Grundlagen der Geometrie*, 7th edition, 1930.
5. ———, *Über die Grundlagen der Geometrie*, Mathematische Annalen, vol. 56, pp. 381-



422. This article is reprinted as appendix IV, pp. 178-230, in the edition of Hilbert's book referred to above.

6. Kerékjártó, *On a geometrical theory of continuous groups*, II. *Euclidean and hyperbolic groups of three dimensional space*, *Annals of Mathematics*, (2), vol. 29, pp. 169-179.

7. Montgomery and Zippin, *Periodic one-parameter groups in three-space*, these *Transactions*, vol. 40 (1936), pp. 24-36.

8. ———, *Compact abelian transformation groups*, *Duke Mathematical Journal*, vol. 4 (1938), pp. 363-373.

9. ———, *Non-abelian compact connected groups of three-space*, *American Journal of Mathematics*, vol. 61 (1939), pp. 375-387.

10. ———, *Topological transformation groups I*, *Annals of Mathematics*, (2), vol. 41 (1940).

11. ———, *A theorem on the rotation group of the two-sphere*, *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 520-521.

12. P. A. Smith, *The topology of transformation groups*, *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 497-514.

13. Veblen and Young, *Projective Geometry*, vol. 2.

SMITH COLLEGE,  
NORTHAMPTON, MASS.,  
QUEENS COLLEGE,  
FLUSHING, N. Y.



## CONFORMALITY IN CONNECTION WITH FUNCTIONS OF TWO COMPLEX VARIABLES

BY

EDWARD KASNER

1. **Introduction.** In the theory of functions of two complex variables  $z = x + iy$ ,  $w = u + iv$ , the transformations of importance are  $Z = Z(z, w)$ ,  $W = W(z, w)$  where  $Z$  and  $W$  are general analytic functions (power series) such that the jacobian  $Z_z W_w - Z_w W_z$  is not identically zero. Any pair of such functions may be regarded as a transformation from the points  $(x, y, u, v)$  to the points  $(X, Y, U, V)$  of a given *real* cartesian four-space  $S_4$ . Poincaré in his fundamental paper in the Palermo Rendiconti (1907) called any such correspondence a *regular* transformation. We employ also the term *pseudo-conformal* transformation. The totality of these transformations forms an infinite group  $G$ . This is *not* the conformal group of the four-space  $S_4$  as is the case for the infinite group of analytic functions  $Z = Z(z)$  of a single complex variable  $z = x + iy$ . As a matter of fact, the theorem of Liouville states that the conformal group of the four-space  $S_4$  is merely the fifteen-parameter group of inversions.

In this paper, we shall obtain several geometric characterizations of this group  $G$  of regular transformations. Our main theorem is that *the group  $G$  of regular (or pseudo-conformal) transformations is characterized by the fact that it leaves invariant the pseudo-angle between any curve  $C$  and any hypersurface  $H$  at their common point of intersection.*

The pseudo-angle may be visualized geometrically as follows. Let a lineal element  $C$  and a hypersurface element  $H$  intersect in a common point  $p$ . Rotate the lineal element  $C$  about the point  $p$  into the hypersurface element  $H$  in the unique planar direction (the isoclinal planar direction), which has the property that the angle between any two lineal elements of the rotation is equal to the angle between their orthogonal projections onto the  $z$ - (and  $w$ -) plane. There is a unique lineal element  $C_1$  in the hypersurface element  $H$ , which is the end result of this rotation. Our pseudo-angle is then the actual angle between the initial lineal element  $C$  and the terminal lineal element  $C_1$  of this rotation.

In conclusion we study Picard's sixteen-parameter group, used in the theory of hyperfuchsian functions. The only pseudo-conformal transformations actually conformal in  $S_4$  constitute a nine-parameter subgroup.

Another geometric interpretation of functions of two complex variables is obtained by using point-pairs (bipoints) in the plane; and this is easily ex-

Presented to the Society, September 11, 1908; also at the Zurich International Congress, 1932; received by the editors May 26, 1939.

tended to  $n$  variables by using  $n$ -points or polygons. See the Bulletin of the American Mathematical Society, vol. 15 (1909), p. 159.

2. **Isoclinal and reverse isoclinal planes.** Before proving these geometric characterizations of the infinite group  $G$ , we shall have to consider some preliminary definitions and theorems. A surface  $s$  of the four-space  $S_4$  is given by the two equations  $F_1(x, y, u, v) = 0$ ,  $F_2(x, y, u, v) = 0$ , where  $F_1$  and  $F_2$  are two independent functions of  $(x, y, u, v)$ . Let  $P_s(x, y, u, v)$  be any point of the surface  $s$ . Construct the orthogonal projections  $P_z(x, y, 0, 0)$  and  $P_w(0, 0, u, v)$  (by means of absolutely perpendicular planes) of the point  $P_s$  on the  $z$ - and  $w$ -planes respectively. Thus the surface  $s$  induces (1) the correspondence  $R_{zw}$  between the points  $P_z$  and  $P_w$  of the  $z$ - and  $w$ -planes, (2) the correspondence  $R_{sz}$  between the points  $P_s$  and  $P_z$  of the  $z$ -plane and the surface  $s$ , and (3) the correspondence  $R_{ws}$  between the points  $P_w$  and  $P_s$  of the  $w$ -plane and the surface  $s$ . We call  $R_{zw}$ ,  $R_{sz}$ ,  $R_{ws}$  the *three correspondences associated with the surface*  $s$ . The two correspondences  $R_{sz}$  and  $R_{ws}$  are the result of orthogonal projections of the points of the surface  $s$  onto the  $z$ - and  $w$ -planes. The correspondence  $R_{zw}$  is given by the equations  $F_1(x, y, u, v) = 0$ ,  $F_2(x, y, u, v) = 0$  of the surface  $s$ . It is noted that any one of these three correspondences may be degenerate.

Since any orthogonal projection of a plane upon a plane in the four-space  $S_4$  preserves parallel lines, we find that for a plane  $\pi$ , each of the three associated correspondences  $R_{zw}$ ,  $R_{sz}$ ,  $R_{ws}$  is an affine transformation. Conversely if any one of the three correspondences  $R_{zw}$ ,  $R_{sz}$ ,  $R_{ws}$  associated with a surface  $s$  is an affine transformation, then all three are affine transformations and the surface  $s$  is a plane. Of course, all of these statements are equivalent to the fact that a plane of the four-space  $S_4$  is given by two independent linear equations in the unknowns  $(x, y, u, v)$ .

For a general plane  $\pi$ , each of the associated correspondences  $R_{zw}$ ,  $R_{sz}$ ,  $R_{ws}$  is an affine transformation. If the associated correspondence  $R_{zw}$  is a direct (or reverse) similitude, then  $\pi$  is termed an *isoclinal plane* (or a *reverse isoclinal plane*). For an isoclinal plane, the correspondences  $R_{sz}$  and  $R_{ws}$  are both direct or reverse similitudes according to the choice of the positive sense of rotation of the angle in  $\pi$ . Similarly for a reverse isoclinal plane  $\pi$ , the correspondences  $R_{sz}$  and  $R_{ws}$  are respectively direct and reverse or reverse and direct similitudes according to the choice of the positive sense of rotation of the angle in  $\pi$ . Thus for an isoclinal or a reverse isoclinal plane, it is found that under each of the three associated correspondences  $R_{zw}$ ,  $R_{sz}$ ,  $R_{ws}$  the angle between any two lines is preserved.

An isoclinal plane may be given by the single complex equation  $w = lz + m$ , where  $l$  and  $m$  are arbitrary complex constants; whereas a reverse isoclinal plane may be given by the single complex equation  $w = l\bar{z} + m$ , where  $\bar{z} = x - iy$  is the conjugate of  $z = x + iy$ . Thus in the totality of  $\infty^6$  planes of the four-space  $S_4$ , there are  $\infty^4$  isoclinal (or reverse isoclinal) planes. These  $\infty^4$  iso-

clinal (or reverse isoclinal) planes form a linear system of planes. Through any given point (or in any hyperplane) of the four-space  $S_4$ , there are  $\infty^2$  isoclinal (or reverse isoclinal) planes. There is one and only one isoclinal (or reverse isoclinal) plane which passes through a given line of the four-space  $S_4$ .

We obtain the following three characterizations of the set of  $2 \times 4$  isoclinal and reverse isoclinal planes among the totality of  $\infty^6$  planes of the four-space  $S_4$ . (1) A plane  $\pi$  is an isoclinal or a reverse isoclinal plane if and only if at least one of the associated affine transformations  $R_{zw}$ ,  $R_{xz}$ ,  $R_{wz}$  is a similitude. (2) The necessary and sufficient condition that a plane  $\pi$  be an isoclinal or a reverse isoclinal plane is that the angle between any line  $L$  of  $\pi$  and its orthogonal projection  $L_z$  (or  $L_w$ ) onto the  $z$ - (or  $w$ -) plane is constant. This result gives the reason for the term isocline. Let  $\phi$  (or  $\psi$ ) be the constant angle between any line  $L$  of the isoclinal or reverse isoclinal plane  $\pi$  and its orthogonal projection  $L_z$  (or  $L_w$ ) onto the  $z$ - (or  $w$ -) plane. Then  $\phi$  and  $\psi$  are complementary angles. (3) A plane  $\pi$  is an isoclinal or a reverse isoclinal plane if and only if the maximum and minimum angles between the plane  $\pi$  and the  $z$ - (or  $w$ -) plane are equal. The common value of the maximum and minimum angles between the isoclinal or reverse isoclinal plane  $\pi$  and the  $z$ - (or  $w$ -) plane is  $\phi$  (or  $\psi$ ). Thus an isoclinal or a reverse isoclinal plane makes complementary angles with the  $z$ - and  $w$ -planes. Also any area in any isoclinal or reverse isoclinal plane is equal to the sum or difference of its orthogonal projections on the  $z$ - and  $w$ -planes. Finally we note that for the isoclinal plane  $w = lz + m$  or the reverse isocline plane  $w = l\bar{z} + m$ , the angle  $\phi$  is  $\arctan |l|$ , where  $|l|$  denotes the absolute value of  $l$ .

**3. Conformal and reverse conformal surfaces.** The envelope of  $\infty^2$  isoclinal (or reverse isoclinal) planes is called a *conformal surface* (or a *reverse conformal surface*). Upon finding the envelope of the  $\infty^2$  isoclinal planes  $w = l(r, t)z + m(r, t)$  (or of the reverse isoclinal planes  $w = l(r, t)\bar{z} + m(r, t)$ ) where  $l$  and  $m$  are complex functions of the *real* variables  $r$  and  $t$ , we find that any conformal (or reverse conformal) surface may be given by the single complex equation  $w = f(z)$  (or  $w = f(\bar{z})$ ), where  $f$  is an analytic function of  $z$  (or  $\bar{z}$ ). From this, it follows that a conformal (or reverse conformal) surface may be given by the two real equations  $u = u(x, y)$ ,  $v = v(x, y)$ , where  $u$  and  $v$  are arbitrary real functions of  $(x, y)$  which satisfy the Cauchy-Riemann equations  $u_x = v_y$ ,  $u_y = -v_x$  (or the reverse Cauchy-Riemann equations  $u_x = -v_y$ ,  $u_y = v_x$ ).

From the above facts, it easily follows that the correspondence  $R_{zw}$  for a conformal (or reverse conformal) surface  $s$  is direct conformal (or reverse conformal). For a conformal surface  $s$ , the correspondences  $R_{zz}$  and  $R_{ww}$  are both direct or reverse conformal transformations according to the choice of the positive sense of rotation of the angle in  $s$ . Similarly for a reverse conformal surface  $s$ , the correspondences  $R_{zz}$  and  $R_{ww}$  are respectively direct and reverse or reverse and direct conformal according to the choice of the positive sense of rotation of the angle in  $s$ . Thus for a conformal (or reverse conformal) sur-

face  $s$ , each of the associated correspondences  $R_{zu}$ ,  $R_{zs}$ ,  $R_{us}$  preserves the angle between two intersecting curves. Conversely if at least one of the associated correspondences  $R_{zu}$ ,  $R_{zs}$ ,  $R_{us}$  of a surface  $s$  is conformal (direct or reverse), then all three are conformal (direct or reverse), and  $s$  is either a conformal or a reverse conformal surface.

**4. Statements of our results.** Under the group  $G$  of regular transformations, every conformal surface is carried into a conformal surface. On the other hand, every reverse conformal surface is *not* carried into a reverse conformal surface. *A transformation  $T$  of the four-space  $S_4$  is regular if and only if it converts every conformal surface into a conformal surface.* The group  $G$  of regular transformations preserves the angle and also the sense of rotation between any two intersecting curves contained in a conformal surface. Thus this group  $G$  induces the group of direct conformal transformations between the conformal surfaces of the four-space  $S_4$ .

If two intersecting curves  $C_1$  and  $C_2$  are tangent to a conformal surface at their common point (or two hypersurfaces  $H_1$  and  $H_2$  intersect in a conformal surface), then under the group  $G$  of regular transformations, the two curves  $C_1$  and  $C_2$  (or the two hypersurfaces  $H_1$  and  $H_2$ ) possess the angle between them as the fundamental differential invariant of the first order. On the other hand, two intersecting curves  $C_1$  and  $C_2$  *not* both tangent to a conformal surface (or two hypersurfaces  $H_1$  and  $H_2$  *not* intersecting in a conformal surface) at their common point do *not* possess any differential invariants of the first order under the infinite group  $G$  of regular transformations. This means that under the group  $G$  of regular transformations, any two concurrent lineal elements not both contained in an isoclinal surface element (or two concurrent hypersurface elements not intersecting in an isoclinal surface element) can be converted into any other two concurrent lineal elements not both contained in an isoclinal surface element (or any other concurrent two hypersurface elements not intersecting in an isoclinal surface element).

The *simplest characterization* of the group  $G$  of regular transformations is connected with the intersection of a curve and a three-dimensional variety. Let a curve  $C$  and a hypersurface  $H$  intersect in a point  $p$ . There is a unique isoclinal plane which passes through the point  $p$  and tangent to the curve  $C$ . Let  $C_1$  be any curve through the point  $p$  which is tangent to this isoclinal plane and to the hypersurface  $H$ . All such curves  $C_1$  are tangent to each other at the point  $p$ . The angle between the curve  $C$  and the curve  $C_1$  is the fundamental differential invariant of the first order between the curve  $C$  and the hypersurface  $H$ . This angle is called the pseudo-angle between the curve  $C$  and the hypersurface  $H$ . *A transformation  $T$  of the four-space  $S_4$  is regular if and only if it preserves the pseudo-angle between any curve  $C$  and any hypersurface  $H$ .* Thus the infinite group  $G$  of regular transformations is characterized by the fact that it leaves invariant the pseudo-angle between every curve  $C$  and every hypersurface  $H$ .

In the final part of our paper, we shall give a brief discussion of the Picard sixteen-parameter group  $G_{16}$  of linear fractional transformations in  $w$  and  $z$ . If a regular transformation  $T$  converts  $4 \infty^2$  isoclinal planes into isoclinal planes, then  $T$  carries every isoclinal plane into an isoclinal plane, and therefore  $T$  is a linear fractional transformation of the group  $G_{16}$ . For any other regular transformation  $T$ , at most  $3 \infty^2$  isoclinal planes become isoclinal planes.

To prove our theorems, we shall have to consider the lineal elements of the four-space  $S_4$  which pass through a given point. Any lineal element through a fixed point may be defined by  $(\rho x', \rho y', \rho u', \rho v')$ , where  $x', y', u', v'$  denote the differentials  $dx, dy, du, dv$  respectively, and  $\rho$  is any real nonzero factor of proportionality. However, to prove our results we shall find it more convenient to define any real lineal element through a given point by the complex coordinates  $(\rho z', \rho w')$ , where  $z' = x' + iy', w' = u' + iv'$ , and  $\rho$  is any real nonzero factor of proportionality.

**5. The necessity of our results.** Let  $T$  be the regular transformation  $Z = Z(z, w)$ ,  $W = W(z, w)$ . Let  $p(x, y, u, v)$  be a fixed point of the four-space  $S_4$  and let  $P(X, Y, U, V)$  be the transformed point under the regular transformation  $T$ . Then the special projective transformation between the two bundles of lineal elements through the points  $p$  and  $P$ , which is induced by the regular transformation  $T$  is given by the equations

$$(1) \quad \rho Z' = \alpha z' + \beta w', \quad \rho W' = \gamma z' + \delta w',$$

where  $\alpha, \beta, \gamma, \delta$  are

$$\begin{aligned} \alpha &= \frac{\partial Z}{\partial z} = \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) (X + iY) = \frac{1}{2} (X_x + Y_y) + \frac{i}{2} (-X_y + Y_x) \\ &= X_x - iX_y, \\ \beta &= \frac{\partial Z}{\partial w} = \frac{1}{2} \left( \frac{\partial}{\partial u} - i \frac{\partial}{\partial v} \right) (X + iY) = \frac{1}{2} (X_u + Y_v) + \frac{i}{2} (-X_v + Y_u) \\ &= X_u - iX_v, \\ (2) \quad \gamma &= \frac{\partial W}{\partial z} = \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) (U + iV) = \frac{1}{2} (U_x + V_y) + \frac{i}{2} (-U_y + V_x) \\ &= U_x - iU_y, \\ \delta &= \frac{\partial W}{\partial w} = \frac{1}{2} \left( \frac{\partial}{\partial u} - i \frac{\partial}{\partial v} \right) (U + iV) = \frac{1}{2} (U_u + V_v) + \frac{i}{2} (-U_v + V_u) \\ &= U_u - iU_v. \end{aligned}$$

Any hypersurface is defined by the equation  $H(x, y, u, v) = 0$ , where  $H$  is any arbitrary real function of  $(x, y, u, v)$ . Thus any hypersurface element through the fixed point  $p(x, y, u, v)$  is given by



$$(3) \quad az' + bw' + \bar{a}\bar{z}' + \bar{b}\bar{w}' = 0,$$

where  $a$  and  $b$  are

$$(4) \quad \begin{aligned} a &= \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) H = \frac{1}{2} (H_x - iH_y), \\ b &= \frac{1}{2} \left( \frac{\partial}{\partial u} - i \frac{\partial}{\partial v} \right) H = \frac{1}{2} (H_u - iH_v). \end{aligned}$$

From (3) and (4), we see that any *real* hypersurface element through the fixed point  $p$  is defined by the complex coordinates  $(\sigma a, \sigma b)$  where  $\sigma$  is a *real* non-zero factor of proportionality.

From (1) and (3), we find that the special projective transformation between the two bundles of hypersurface elements through the fixed points  $p$  and  $P$ , which is induced by the regular transformation  $T$ , is given by

$$(5) \quad \sigma a = \alpha A + \gamma B, \quad \sigma b = \beta A + \delta B.$$

Since the equation of any conformal surface is of the form  $w=f(z)$  where  $f$  is an analytic function of  $z$ , there follows from the equations of any regular transformation  $T$

**THEOREM 1.** *Under the group  $G$  of regular transformations, every conformal surface is converted into a conformal surface.*

Since every conformal surface becomes a conformal surface, it follows that under the group  $G$  of regular transformations, every isoclinal surface element is carried into an isoclinal surface element. This is also a consequence of equations (1) upon observing that the equation of any isoclinal surface element through the fixed point  $p$  is  $w' = lz'$ , where  $l$  is an arbitrary complex constant.

Two lineal elements are said to be an isoclinal pair if they are contained in an isoclinal surface element. The condition for an isoclinal pair of lineal elements is

$$(6) \quad \frac{z_2'}{z_1'} = \frac{w_2'}{w_1'} = \text{complex constant (not real)}.$$

Two hypersurface elements are said to form an isoclinal pair if they intersect in an isoclinal surface element. The condition for an isoclinal pair of hypersurface elements is

$$(7) \quad \frac{a_2}{a_1} = \frac{b_2}{b_1} = \text{complex constant (not real)}.$$

From equations (1) and (6), we obtain



**THEOREM 2.** *Two intersecting curves  $C_1$  and  $C_2$  which are tangent to a conformal surface at their common point possess the fundamental differential invariant of first order*

$$(8) \quad \text{amp} \frac{z'_2}{z'_1} = \text{amp} \frac{w'_2}{w'_1}.$$

*This is the angle between the two curves  $C_1$  and  $C_2$ . It can be written in the real form*

$$(9) \quad \text{arc tan} \frac{dx_1 dy_2 - dx_2 dy_1}{dx_1 dx_2 + dy_1 dy_2} = \text{arc tan} \frac{du_1 dv_2 - du_2 dv_1}{du_1 du_2 + dv_1 dv_2}.$$

By equations (5) and (7), we obtain the following dual result:

**THEOREM 3.** *Two hypersurfaces  $H_1$  and  $H_2$  which intersect in a conformal surface possess the fundamental differential invariant of first order*

$$(10) \quad \text{amp} \frac{a_2}{a_1} = \text{amp} \frac{b_2}{b_1}.$$

*This is the angle between the two hypersurfaces  $H_1$  and  $H_2$ . It can be written in the real form*

$$(11) \quad \text{arc tan} \frac{H_{1x} H_{2y} - H_{1y} H_{2x}}{H_{1x} H_{2x} + H_{1y} H_{2y}} = \text{arc tan} \frac{H_{1u} H_{2v} - H_{1v} H_{2u}}{H_{1u} H_{2u} + H_{1v} H_{2v}}.$$

Let us now consider the case where two intersecting curves  $C_1$  and  $C_2$  are not both tangent to a conformal surface at their common point. In that case, we can convert any non-isoclinal pair of lineal elements  $(\rho_1 Z'_1, \rho_1 W'_1)$  and  $(\rho_2 Z'_2, \rho_2 W'_2)$  into the lineal elements  $(1, 0)$  and  $(0, 1)$ , which of course are a non-isoclinal pair of lineal elements. The most general transformation of form (1) that will do this is

$$(12) \quad \rho Z' = \rho_1 Z'_1 z' + \rho_2 Z'_2 w', \quad \rho W' = \rho_1 W'_1 z' + \rho_2 W'_2 w'.$$

This is an admissible transformation since the jacobian  $J = \rho_1 \rho_2 (Z'_1 W'_2 - Z'_2 W'_1)$  is not zero. Hence we have proved that two intersecting curves  $C_1$  and  $C_2$  not tangent to a conformal surface at their common point have no differential invariants of the first order. The dual results for hypersurfaces are also valid. Thus we have

**THEOREM 4.** *Two intersecting curves  $C_1$  and  $C_2$  not both tangent to a conformal surface at their common point (or two hypersurfaces  $H_1$  and  $H_2$  which do not intersect in a conformal surface) possess no differential invariants of the first order.*

Let  $C(z', w')$  be a given lineal element and  $H(a, b)$  a given hypersurface

element. There is a unique isoclinal surface element which contains the curve  $C(z', w')$ . It is given by

$$(13) \quad \frac{Z'}{z'} = \frac{W'}{w'} = \lambda,$$

where  $\lambda$  is a complex constant (not real). Upon substituting this into the equation  $aZ' + bW' + \bar{a}\bar{Z}' + \bar{b}\bar{W}' = 0$  of the hypersurface element  $H(a, b)$ , we find that the lineal element  $C_1$  of intersection between the isoclinal surface element (13) and the hypersurface element  $H(a, b)$  is given by the equation

$$(14) \quad \frac{Z'}{z'} = \frac{W'}{w'} = i(\bar{a}\bar{z}' + \bar{b}\bar{w}').$$

Since, according to Theorem 2, the angle between the curves  $C$  and  $C_1$  is invariant, we obtain

**THEOREM 5.** *A curve  $C$  and a hypersurface  $H$  which intersect in a common point possess the fundamental differential invariant of first order*

$$(15) \quad \frac{1}{2}\pi - \text{amp}(az' + bw'),$$

*evaluated at the common point. This is called the pseudo-angle between the curve  $C$  and the hypersurface  $H$ . The pseudo-angle represents the angle between the curve  $C$  and any curve  $C_1$  through the point  $p$  such that  $C$  and  $C_1$  are tangent to a conformal surface at the point  $p$ , and  $C_1$  is tangent to the hypersurface  $H$  at the point  $p$ . Dually, we find that the pseudo-angle represents the angle between the hypersurface  $H$  and any hypersurface  $H_1$  through the point  $p$  such that  $H$  and  $H_1$  intersect in a conformal surface and  $H_1$  is tangent to the curve  $C$  at the point  $p$ . This pseudo-angle can be written in the real form*

$$(16) \quad \text{arc tan} \frac{H_z dx + H_y dy + H_u du + H_v dv}{-H_v dx + H_z dy - H_u du + H_v dv}.$$

The fact that this is the only differential invariant of the first order between a curve  $C$  and a hypersurface  $H$  which pass through a given point  $p$  is an immediate consequence of equations (1) and (5).

**6. The sufficiency of our results.** Let a general transformation  $T$ ,

$$(17) \quad X = (x, y, u, v), \quad Y = Y(x, y, u, v), \quad U = U(x, y, u, v), \quad V = V(x, y, u, v),$$

be given.  $T$  is not necessarily a regular transformation. Let  $p(x, y, u, v)$  be a fixed point of the four-space  $S_4$  and let  $P(X, Y, U, V)$  be the transformed point under the transformation  $T$ . Then  $T$  induces the following general projective transformation between the two bundles of lineal elements through the points  $p$  and  $P$ :

$$\begin{aligned}
 (18) \quad \rho X' &= X_z x' + X_y y' + X_u u' + X_v v', \\
 \rho Y' &= Y_z x' + Y_y y' + Y_u u' + Y_v v', \\
 \rho U' &= U_z x' + U_y y' + U_u u' + U_v v', \\
 \rho V' &= V_z x' + V_y y' + V_u u' + V_v v'.
 \end{aligned}$$

Changing (18) from the real notation  $(x', y', u', v')$  to the complex notation by means of the equations

$$\begin{aligned}
 (19) \quad Z' &= X' + iY', & x' &= \frac{1}{2}(z' + \bar{z}'), & y' &= \frac{i}{2}(\bar{z}' - z'), \\
 W' &= U' + iV', & u' &= \frac{1}{2}(w' + \bar{w}'), & v' &= \frac{i}{2}(\bar{w}' - w'),
 \end{aligned}$$

we find that equations (18) may be written in the compact complex form

$$(20) \quad \rho Z' = \alpha z' + \beta w' + \phi \bar{z}' + \psi \bar{w}', \quad \rho W' = \gamma z' + \delta w' + \chi \bar{z}' + \omega \bar{w}',$$

where  $\alpha, \beta, \gamma, \delta, \phi, \psi, \chi, \omega$  are given:

$$\begin{aligned}
 (21) \quad \alpha &= \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) (X + iY) = \frac{1}{2} (X_z + Y_y) + \frac{i}{2} (-X_y + Y_z), \\
 \beta &= \frac{1}{2} \left( \frac{\partial}{\partial u} - i \frac{\partial}{\partial v} \right) (X + iY) = \frac{1}{2} (X_u + Y_v) + \frac{i}{2} (-X_v + Y_u), \\
 \gamma &= \frac{1}{2} \left( \frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) (U + iV) = \frac{1}{2} (U_z + V_y) + \frac{i}{2} (-U_y + V_z), \\
 \delta &= \frac{1}{2} \left( \frac{\partial}{\partial u} - i \frac{\partial}{\partial v} \right) (U + iV) = \frac{1}{2} (U_u + V_v) + \frac{i}{2} (-U_v + V_u), \\
 \phi &= \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) (X + iY) = \frac{1}{2} (X_z - Y_y) + \frac{i}{2} (X_y + Y_z), \\
 \psi &= \frac{1}{2} \left( \frac{\partial}{\partial u} + i \frac{\partial}{\partial v} \right) (X + iY) = \frac{1}{2} (X_u - Y_v) + \frac{i}{2} (X_v + Y_u), \\
 \chi &= \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) (U + iV) = \frac{1}{2} (U_z - V_y) + \frac{i}{2} (U_y + V_z), \\
 \omega &= \frac{1}{2} \left( \frac{\partial}{\partial u} + i \frac{\partial}{\partial v} \right) (U + iV) = \frac{1}{2} (U_u - V_v) + \frac{i}{2} (U_v + V_u).
 \end{aligned}$$

The transformation (20) is thus the general projective transformation (18) between the two bundles of lineal elements through the two points  $p$  and  $P$ .

Let the transformation  $T$  carry every conformal surface into a conformal

surface. Then  $T$  must convert every isoclinal surface element into an isoclinal surface element. Hence (20) must carry every equation of the form  $w' = lz'$  into an equation of the same form. For this to be so, we must have

$$(22) \quad \phi = \psi = \chi = \omega = 0.$$

These are the double Cauchy-Riemann equations for the two complex functions  $X+iY$  and  $U+iV$ . Hence these functions must be analytic functions of  $z$  and  $w$ . Thus

**THEOREM 6.** *Any transformation  $T$  of the four-space  $S_4$  which converts every conformal surface into a conformal surface is a regular transformation. Thus the infinite group  $G$  of regular transformations is characterized by the fact that it preserves conformal surfaces.*

Next we shall prove that the pseudo-angle (the differential invariant (15) or (16)) of Theorem 5 characterizes the infinite group  $G$  of regular transformations. Let the transformation  $T$  preserve the differential invariant (15) between every lineal element  $c(z', w')$  and every hypersurface element  $h(a, b)$  which passes through the common point  $p$ . Then under  $T$  we must have

$$(23) \quad \frac{az' + bw'}{\bar{a}\bar{z}' + \bar{b}\bar{w}'} = \frac{AZ' + BW'}{\bar{A}\bar{Z}' + \bar{B}\bar{W}'},$$

where the capital letters denote the transformed lineal element  $C(Z', W')$  and the transformed hypersurface element  $H(A, B)$ .

First we shall show that any isoclinal pair of lineal elements  $c_1(z'_1, w'_1)$  and  $c_2(z'_2, w'_2)$  is converted into an isoclinal pair of lineal elements  $C_1(Z'_1, W'_1)$  and  $C_2(Z'_2, W'_2)$ . Since  $c_1(z'_1, w'_1)$  and  $c_2(z'_2, w'_2)$  are contained in an isoclinal surface element, we must have

$$(24) \quad \frac{z'_2}{z'_1} = \frac{w'_2}{w'_1} = \lambda,$$

where  $\lambda$  is a fixed non-real complex number. Let us pass any one of the  $\infty^2$  hypersurface elements  $h(a, b)$  through the lineal element  $c_1(z'_1, w'_1)$ . Then under  $T$  the transformed hypersurface element  $H(A, B)$  must contain the transformed lineal element  $C_1(Z'_1, W'_1)$ . Hence we must have

$$(25) \quad az'_1 + bw'_1 + \bar{a}\bar{z}'_1 + \bar{b}\bar{w}'_1 = 0, \quad AZ'_1 + BW'_1 + \bar{A}\bar{Z}'_1 + \bar{B}\bar{W}'_1 = 0.$$

Under the transformation  $T$ , the pseudo-angle between the lineal element  $c_2(z'_2, w'_2)$  and any one of the  $\infty^2$  hypersurface elements  $h(a, b)$  through the lineal element  $c_1(z'_1, w'_1)$  must be equal to the pseudo-angle between the transformed lineal element  $C_2(Z'_2, W'_2)$  and the corresponding transformed hypersurface element  $H(A, B)$ . This means that the equation (23) must be valid for these lineal and hypersurface elements. Then because of (24) and

(25), the equation (23) becomes

$$(26) \quad \frac{\lambda}{\bar{\lambda}} = \frac{AZ'_2 + BW'_2}{\overline{AZ'_2} + \overline{BW'_2}}.$$

This equation must be true for all the  $\infty^2$  hypersurface elements  $H(A, B)$  which pass through the lineal element  $C_1(Z'_1, W'_1)$ .

From this equation, and from the fact that  $AZ'_1 + BW'_1 = -(\overline{AZ'_1} + \overline{BW'_1})$ ,

$$(27) \quad AZ'_2 + BW'_2 = i\rho_1\lambda, \quad \overline{AZ'_1} + \overline{BW'_1} = i\rho_2,$$

where  $\rho_1$  and  $\rho_2$  are arbitrary real numbers. Let us suppose that  $Z'_2/Z'_1 \neq W'_2/W'_1$ . From these two equations, we can solve for  $A$  and  $B$  in terms of the arbitrary real numbers  $\rho_1$  and  $\rho_2$ . Thence  $A$  and  $B$  are linear homogeneous functions of  $\rho_1$  and  $\rho_2$ . This proves that the equation (26) can hold for only  $\infty^1$  hypersurface elements passing through the lineal element  $C_1(Z'_1, W'_1)$ . This contradicts the fact that the equation (26) must hold for all the hypersurface elements  $H(A, B)$  through the lineal element  $C_1(Z'_1, W'_1)$ . Hence we must have

$$(28) \quad \frac{Z'_2}{Z'_1} = \frac{W'_2}{W'_1}.$$

This shows that the transformed lineal elements  $C_1(Z'_1, W'_1)$  and  $C_2(Z'_2, W'_2)$  must be contained in an isoclinal surface element. Therefore every isoclinal pair of lineal elements is converted by  $T$  into an isoclinal pair of lineal elements.

Since any isoclinal pair of lineal elements is carried by  $T$  into an isoclinal pair of lineal elements, it follows that  $T$  carries every isoclinal surface element into an isoclinal surface element. Hence every conformal surface becomes a conformal surface and the transformation  $T$  must therefore be a regular transformation. Thus we have proved

**THEOREM 7.** *Any transformation  $T$  of the four-space  $S_4$  which preserves the pseudo-angle (the differential expression of the first order (15) or (16)) between every curve and every hypersurface evaluated at their common point must be a regular transformation. Thus the infinite group  $G$  of regular transformations is characterized by the fact that it leaves invariant the pseudo-angle between every curve and every hypersurface.*

**7. The Picard sixteen-parameter group  $G_{16}$  of linear fractional transformations.** In this section, we shall give a characterization of the group  $G_{16}$  of the linear fractional transformations in  $z$  and  $w$

$$(29) \quad Z = \frac{a_1z + b_1w + c_1}{az + bw + c}, \quad W = \frac{a_2z + b_2w + c_2}{az + bw + c}.$$

Any transformation of the form (29) is a quadric Cremona transformation. It may be considered to be a direct generalization of the Moebius group of circular transformations. Of course, it is *not* the inversion group of the four-space  $S_4$ . As a matter of fact, any hypersphere (or any hyperplane) is converted by (29) into a special type of quadric hypersurface.

Under any regular transformation  $T$ , let us find what isoclinal planes become isoclinal planes. For this to be so, the differential equation  $d^2w/dz^2=0$  must be carried into the differential equation  $d^2W/dZ^2=0$ . Hence those isoclinal planes which become isoclinal equations under the regular transformation  $T$  must satisfy the equation

$$(30) \quad \left( Z_z + \frac{dw}{dz} Z_w \right) \left[ W_{zz} + 2 \frac{dw}{dz} W_{zw} + \left( \frac{dw}{dz} \right)^2 W_{ww} \right] \\ - \left( W_z + \frac{dw}{dz} W_w \right) \left[ Z_{zz} + 2 \frac{dw}{dz} Z_{zw} + \left( \frac{dw}{dz} \right)^2 Z_{ww} \right] = 0.$$

First, if this equation is an identity in  $dw/dz$ , we find that  $Z$  and  $W$  must be given by the equations (29). That is, the group  $G_{16}$  of linear fractional transformations as given by the equations (29) convert every isoclinal plane into an isoclinal plane.

Next if the above equation is not identically zero, we can solve (30) for  $dw/dz$  and obtain at most three differential equations of the form

$$(31) \quad \frac{dw}{dz} = f(z, w),$$

where  $f$  is an analytic function of  $z$  and  $w$ . Any such differential equation contains  $\infty^2$  solutions. Thus we have proved

**THEOREM 8.** *If a regular transformation  $T$  converts  $4 \infty^2$  isoclinal planes into isoclinal planes, then every isoclinal plane is converted into an isoclinal plane, and therefore  $T$  is a transformation of the group  $G_{16}$  of the linear fractional transformations as given by equations (29). Any other regular transformation  $T$  converts at most  $3 \infty^2$  isoclinal planes into isoclinal planes.*

It is found that, under the group  $G_{16}$  of fractional linear transformations as given by (29), the family of quadric hypersurfaces

$$(32) \quad az\bar{z} + bw\bar{w} + \gamma zw + \bar{\gamma}z\bar{w} + \delta z + \epsilon w + \bar{\delta}\bar{z} + \bar{\epsilon}\bar{w} + f = 0,$$

where  $a, b, f$  are arbitrary real constants and  $\gamma, \delta, \epsilon$  are arbitrary complex constants, is converted into itself. The real form of this family of quadric hypersurfaces is

$$(33) \quad a(x^2 + y^2) + b(u^2 + v^2) + 2c_1(ux + vy) + 2c_2(-uy + vx) \\ + 2d_1x + 2d_2y + 2e_1u + 2e_2v + f = 0.$$



There are  $\infty^8$  hypersurfaces in this family. Every hypersphere (or every hyperplane) of the four-space  $S_4$  becomes a special quadric hypersurface of the form (32) or (33). Also the intersection of any isoclinal plane with this special quadric hypersurface is a circle. Thus any transformation of the form (29) induces a Moebius circular transformation between the isoclinal planes of the four-space  $S_4$ . In this respect, the group  $G_{16}$  of linear fractional transformations in  $z$  and  $w$  may be regarded as a generalization of the Moebius group of circular transformations to four-space. Also the family of special quadric hypersurfaces (32) or (33) can be considered to be a generalization of the family of circles.

In conclusion I wish to express my thanks to Dr. J. De Cicco for his valuable assistance in writing this paper.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

## ARC- AND TREE-PRESERVING TRANSFORMATIONS

BY

D. W. HALL<sup>(1)</sup> AND G. T. WHYBURN

1. **Introduction.** In an earlier paper<sup>(2)</sup> by one of us, referred to hereafter as A.P.T., arc-preserving transformations were defined and studied in connection with an irreducibility condition on the transformation. It was shown, for example, that if  $A$  and  $B$  are compact locally connected metric continua which are cyclic (that is, without cut points) any single valued continuous arc-preserving and irreducible transformation  $T(A)=B$  of  $A$  onto  $B$  is necessarily a homeomorphism. ("Arc-preserving" means that the image of every simple arc in  $A$  is either a simple arc or a single point of  $B$ ; irreducibility of  $T$  means that no proper subcontinuum of  $A$  maps onto all of  $B$ .) It was shown, furthermore, that in case  $A$  is hereditarily locally connected the same conclusion holds without the assumption of irreducibility; and the prediction was made that this is true in the general case.

Now as pointed out in A.P.T., if  $A$  is a compact continuum and  $T(A)=B$  is continuous, then, since the property of being a subcontinuum of  $A$  mapping onto all of  $B$  under  $T$  is inducible, there always exists a subcontinuum  $A_1$  of  $A$  such that  $T(A_1)=B$  and  $T$  is irreducible on  $A_1$ . However, since local connectedness of  $A$  would certainly not in general insure local connectedness of  $A_1$ , it is not possible always to reduce the set  $A$  so as to make the transformation irreducible without sacrificing essential properties of  $A$ .

In the present paper we propose not only to completely justify the earlier prediction referred to above, but also to obtain theorems concerning a much more general type of transformation than "arc-preserving" which will give all the theorems of the first three sections of A.P.T. as immediate corollaries.

R. G. Simond<sup>(3)</sup> has studied tree-preserving transformations on locally connected compact and metric continua (that is, transformations  $T(A)=B$  satisfying the condition that the image of every tree (or dendrite) in  $A$  is a tree in  $B$ ). Miss Simond has proved with considerable difficulty that every arc-preserving transformation is tree-preserving. We show that this result is an immediate consequence of one of our theorems, as it is also of a theorem of A.P.T. In fact our first principal result, the proof of which is very much simpler than that given by Simond, shows that in order that  $T(A)=B$  be tree-preserving it is necessary and sufficient that the image of every simple

---

Presented to the Society, December 28, 1939; received by the editors January 8, 1940.

<sup>(1)</sup> This work was started when the first named author was a National Research Fellow at the University of Virginia.

<sup>(2)</sup> See G. T. Whyburn, *Arc-preserving transformations*, American Journal of Mathematics, vol. 58 (1936), pp. 305-312.

<sup>(3)</sup> Duke Mathematical Journal, vol. 4 (1938), pp. 575-589.

arc in  $A$  shall be a tree in  $B$ . We also give an independent proof of the Simond theorem.

If  $A$  is a tree, every simple arc in  $A$  is a cyclic chain<sup>(4)</sup> in  $A$ . This shows at once that if  $A$  and  $B$  are both trees and  $T(A)=B$  is continuous, then in order that  $T$  be arc-preserving it is necessary and sufficient that  $T$  be cyclic chain-preserving. This immediately suggests another of our main results: If  $A$  is a compact locally connected continuum and  $T(A)=B$  is continuous, then in order that  $T$  be arc-preserving it is necessary and sufficient that  $T$  be both tree-preserving and cyclic chain-preserving. We give this theorem added meaning by obtaining a somewhat unexpected characterization of tree-preserving transformations in terms of the action of these transformations and their inverses on the sets  $A$  and  $B$ .

In §5 a characterization of  $A$ -set reversing transformations<sup>(5)</sup> is given which supplements the treatment of this type of transformation initiated in A.P.T.

In conclusion we might mention that if  $B$  is cyclic, the following types of transformations are equivalent: (a) arc-preserving, (b) tree-preserving, (c)  $A$ -set reversing, (d) monotone retracting.

Throughout the paper all transformations are assumed to be single valued and continuous and all continua compact and metric.

**2. Principal results.** We assume throughout this section that  $A$  and  $B$  are locally connected continua and that  $T(A)=B$  maps every arc of  $A$  onto a dendrite in  $B$ .

(2.0) *The image of every dendritic graph in  $A$  is a dendrite in  $B$ .*

**Proof.** (By induction on the number of end points.) If the number of end points in the graph is 2, then the hypothesis gives the conclusion since the graph is an arc. Suppose that any dendrite in  $A$  having  $k$  or less end points ( $k > 0$ ) maps onto a dendrite in  $B$ . Let  $D$  be a dendrite in  $A$  having  $k+1$  end points. Let  $p$  be a branch point in  $D$  giving a decomposition  $D=D_1+D_2+D_3$ , where  $D_1$ ,  $D_2$ , and  $D_3$  are dendrites intersecting by pairs in  $p$ . Since  $D_1$ ,  $D_2$ ,  $D_3$ ,  $D_1+D_2$ ,  $D_1+D_3$ ,  $D_2+D_3$  are dendrites having at most  $k$  end points, each of their transforms is a tree. Hence  $T(D_1) \cdot T(D_2)$  and  $T(D_2) \cdot T(D_3)$  are con-

<sup>(4)</sup> If  $M$  is a locally connected compact and metric continuum and  $A$  is a closed subset of  $M$  containing every simple arc  $axb$  of  $M$  such that  $a$  and  $b$  are points of  $A$ , then  $A$  is called an  $A$ -set. By the cyclic chain in  $M$  determined by two points  $a$  and  $b$  of  $M$  and designated by  $C(a, b)$  is meant the product of all  $A$ -sets in  $M$  containing both  $a$  and  $b$ . It is the minimal  $A$ -set in  $M$  containing both these points. The cyclic chains in  $M$  are closely related to the decomposition of  $M$  into its cyclic elements, for which see Kuratowski and Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331. See also G. T. Whyburn, *American Journal of Mathematics*, vol. 50 (1928), pp. 167-194, and W. L. Ayres, *these Transactions*, vol. 30 (1928), pp. 567-578, and vol. 31 (1929), pp. 595-695.

<sup>(5)</sup> For definition, see §5.

nected; and since both of these sets contain  $T(p)$ , their sum  $T(D_3) \cdot T(D_1 + D_2)$  is connected. Thus  $T(D) = T(D_1 + D_2 + D_3)$  is a dendrite.

(2.1) *If  $B$  is cyclic,  $T$  is  $A$ -set reversing.*

**Proof.** Otherwise there exists a simple arc  $b_1xb_2$  in  $A$  such that  $T(b_1) = T(b_2) = b$ , but  $T^{-1}(b) \cdot b_1xb_2 = b_1 + b_2$ . Let  $T(b_1xb_2) = X$  and  $K = T^{-1}(X)$ . From (2.0) it follows that  $X$  is a dendrite; hence  $X \neq B$ . Thus  $A - K \neq \emptyset$ .

(i) *For any component  $R$  of  $A - K$ ,  $T(F(R))$  is a single point<sup>(6)</sup>.*

Otherwise there exists a simple arc  $cyd$  in  $R + c + d$  such that  $c$  and  $d$  lie in  $F(R)$  and  $T(c) \neq T(d)$ . But now if both  $c$  and  $d$  are on  $b'xb''$  (in the order  $b', c, d, b''$ ) we let  $t = b'c + cyd + db''$ ; if not let  $t$  be a dendritic graph in  $A$  which contains both  $b_1xb_2$  and  $cyd$ . This is impossible by (2.0) since in either case  $T(t)$  must contain a simple closed curve.

(ii) *There exist two components  $R$  and  $S$  of  $A - K$  such that  $a = T(F(R)) \neq T(F(S)) = c$  and  $T(R) \cdot T(S) \neq \emptyset$ .*

For let  $R$  be any component of  $A - K$  and let  $Q$  be the sum of all those components  $U$  of  $A - K$  such that  $T(F(U)) = T(F(R)) = a$ . Then since  $Q + T^{-1}(a)$  is closed, it follows that  $T(Q) + a$  is closed. Since  $a$  is not a cut point of  $B$ , it follows that some point  $x$  of  $T(Q)$  must be a limit point of  $B - T(Q)$ . (We know that  $T(Q) + a \neq B$  since  $X$  is not a single point.) But  $Q$  is open in  $A$ ; consequently  $T^{-1}(x)$  must intersect some component  $S$  of  $A - K$  which does not belong to  $Q$ . Thus if we set  $c = T(F(S))$ , (ii) is satisfied.

Now to prove (2.1), let  $y'$  and  $y''$  be points of  $R$  and  $S$ , respectively, such that  $T(y') = T(y'') = y$ . Then there exists a dendritic graph  $t'$  in  $A$  containing  $y'$  and  $b_1xb_2$ . If  $t'$  contains  $y''$ , let  $t = t'$ . Otherwise there exists a dendritic graph  $t$  in  $A$  containing both  $y''$  and  $t'$ . It is immediate that  $T(t)$  contains a simple closed curve, contradicting (2.0).

We have at once:

(2.11) *If  $A$  and  $B$  are both cyclic,  $T$  is a homeomorphism.*

(2.2) *In order that a single valued continuous transformation  $T(A) = B$  shall be tree-preserving it is necessary and sufficient that the image of every arc in  $A$  shall be a tree in  $B$ .*

**Proof.** The necessity is trivial. To prove the sufficiency suppose  $A$  is a tree and that  $B$  has a true cyclic element  $E_b$ . Let  $W(B) = E_b$  be monotone and retracting. Then  $WT(A) = E_b$  is a transformation which maps arcs into trees. Thus, by (2.1),  $WT$  is  $A$ -set reversing, hence monotone. But this makes  $E_b$  a tree, which is absurd.

<sup>(6)</sup> For any open set  $G$ ,  $F(G)$  denotes the boundary of  $G$ , that is, the set  $\bar{G} - G$ .

(2.3) *If  $B$  is cyclic and no  $A$ -set in  $A$  other than  $A$  itself maps onto all of  $B$ , then  $A$  is cyclic and  $T$  is a homeomorphism.*

**Proof.** For if  $A$  had a cut point  $p$ , we could write  $A = A_1 + A_2$ , where  $A_1$  and  $A_2$  are  $A$ -sets with  $A_1 \cdot A_2 = p$ . Then since, by (2.1),  $T$  is monotone, either  $A_1 - p$  or  $A_2 - p$ , say  $A_1 - p$ , contains the set  $T^{-1}(B - T(p))$ , as this latter set is connected. But this gives  $T(A_1) = B$ .

(2.4) *If  $B$  is cyclic,  $T$  is equivalent to a monotone transformation retracting  $A$  onto some true cyclic element of  $A$ . Thus, in this situation, "arc-preserving," "tree-preserving," " $A$ -set reversing," and "monotone retracting" are all equivalent.*

**Proof.** For let  $E_a$  be a minimal  $A$ -set in  $A$  mapping onto all of  $B$  under  $T$ . Then since  $T(E_a) = B$  is a transformation mapping arcs into trees, it follows from (2.3) that  $E_a$  is a cyclic element of  $A$  and  $E_a$  maps onto all of  $B$  topologically under  $T$ . Thus if for each  $y$  in  $B$  we set  $h(y) = E_a \cdot T^{-1}(y)$ , then  $h$  is topological and the transformation  $hT(A) = E_a$  is retracting. Furthermore,  $hT$  is monotone since both  $h$  and  $T$  are monotone. Obviously  $hT$  is equivalent to  $T$ , since  $h^{-1}(hT) = T$ .

**DEFINITION.** For any true cyclic element  $E_a$  of  $A$  we define  $E_a^0$  as the set of all internal points of  $E_a$ , that is, all points  $x$  of  $E_a$  which are non-cut points of  $A$ . It is well known that the set of all non-internal points of any such  $E_a$  is countable.

(2.5) *For each true cyclic element  $E_b$  of  $B$  there exists a unique true cyclic element  $E_a$  of  $A$  such that  $T(E_a) = E_b$ . The transformation  $T$  is a homeomorphism on  $E_a$  and  $T^{-1}$  is single valued on  $T(E_a^0)$ .*

**Proof.** For let  $W(B) = E_b$  be monotone and retracting. Since  $WT(A) = E_b$  is a transformation which maps arcs into trees, it follows from (2.4) that there exists a true cyclic element  $E_a$  of  $A$  and a homeomorphism  $h(E_b) = E_a$  such that  $hWT(A) = E_a$  is monotone and retracting. Since  $E_a$  maps topologically under  $hWT$ , it must therefore map topologically under  $T$ . Let  $y = T(x)$  be a point of  $T(E_a^0)$ , where  $x$  lies in  $E_a^0$ . Since  $x$  is an internal point of  $E_a$  and  $hWT$  is monotone and retracting, we see that  $x = (hWT)^{-1}(y) = T^{-1}W^{-1}h^{-1}(y) = T^{-1}W^{-1}(y) \supset T^{-1}(y)$ , and hence  $x = T^{-1}(y)$ . The uniqueness of  $E_a$  follows at once from this single-valuedness of  $T^{-1}$ .

(2.6) *Let  $A$  be a compact locally connected continuum and  $T(A) = B$  be continuous. Then in order that  $T$  be tree-preserving it is necessary and sufficient that for each true cyclic element  $E_b$  of  $B$  there exist a true cyclic element  $E_a$  of  $A$  mapping onto  $E_b$  topologically under  $T$  and such that  $T^{-1}$  is single valued on the set  $T(E_a^0)$ .*

**Proof.** The necessity follows from (2.5). To establish the sufficiency we need only, in virtue of (2.2), show that the image of every simple arc  $t$  in  $A$



is a tree in  $B$ . Assuming the contrary,  $T(t)$  must contain a simple closed curve  $J'$ . Let  $E_b$  be the true cyclic element of  $B$  containing  $J'$  and  $E_a$  the true cyclic element of  $A$  which satisfies the conditions of the theorem. It follows at once that  $t \cdot E_a$  is a simple arc  $axb$  which maps into an arc  $a'x'b'$  of  $J'$ . Let  $a'y'b'$  be the other arc of  $J'$ , and suppose that  $y'$  is the image of an internal point of  $E_a$ . Then  $T^{-1}(y')$  contains a point of  $E_a$  and a point of  $t - E_a$ , which is impossible.

(2.7) *Let  $A$  be a compact locally connected continuum and let  $T(A) = B$  be continuous. Then in order that  $T$  be arc-preserving it is necessary and sufficient that it be tree-preserving and that the image of each cyclic chain<sup>(7)</sup> in  $A$  be a cyclic chain in  $B$ .*

**Proof. Necessity:** The first condition is necessary by (2.2). That the second condition is necessary results essentially from the fact that, for arc-preserving transformations, (1)  $A$ -sets map onto  $A$ -sets, and (2) the property of having any three points on an arc is invariant.

We first show that  $A$ -sets map onto  $A$ -sets. Let  $A'$  be an  $A$ -set in  $A$  and  $T(A') = B'$ . For any cyclic element  $E_b$  of  $B$  intersecting  $B'$  in at least two points, let  $E_a$  be the corresponding cyclic element of  $A$  given by (2.6). Since  $E_b \cdot B'$  is a nondegenerate continuum and  $T(E_a) = E_b$  is topological,  $E_b \cdot B'$  must contain the image  $y$  of at least one internal point  $x$  of  $E_a$ . Then since  $x = T^{-1}(y)$  we see that  $x$  must lie in  $A'$ . Thus  $E_a$  is contained in  $A'$ , consequently  $E_b$  is contained in  $B'$ . Therefore,  $B'$  is an  $A$ -set in  $B$ .

Now to prove the necessity of the second condition of the theorem, let  $C(a, b)$  be a cyclic chain in  $A$ . Then  $T(C(a, b)) = K$  is an  $A$ -set in  $B$ . Let  $x, y, z$  be points of  $K$  and  $x', y', z'$  be points of  $C(a, b) \cdot T^{-1}(x)$ ,  $C(a, b) \cdot T^{-1}(y)$ ,  $C(a, b) \cdot T^{-1}(z)$ , respectively. There exists an arc  $cd$  in  $C(a, b)$  containing  $x', y'$ , and  $z'$ . Hence  $T(cd)$  is an arc in  $K$  containing  $x, y$ , and  $z$ . Therefore,  $K$  is a cyclic chain (since for  $A$ -sets the property of being a cyclic chain is equivalent to the property of containing an arc through any three points).

**Sufficiency:** Let  $ab$  be any simple arc in  $A$ . We first show that if  $E_b$  is any true cyclic element of  $B$  such that  $E_b \cdot T(ab)$  is nondegenerate and  $E_a$  is the corresponding cyclic element of  $A$  given by (2.6) and  $xy$  is the arc  $E_a \cdot ab$ , then  $T(ax + yb) \cdot E_b = x + y$ . (We may suppose the order  $a, x, y, b$ .) If this is not so, then  $T(ax) \cdot E_b$  or  $T(yb) \cdot E_b$ , say  $T(ax) \cdot E_b$ , is a nondegenerate continuum; hence there must exist a point  $z$  distinct from  $T(x)$  of  $T(E_a^0)$  which belongs to  $E_b \cdot T(ax)$ . This contradicts (2.6), since  $T^{-1}(z)$  intersects both  $E_a$  and  $ax - x$ . Thus  $E_b \cdot T(ab)$  is a simple arc  $x'y' = T(xy)$ . Furthermore, no interior point of  $x'y'$  is a limit point of  $T(ab) - x'y'$ .

Now if  $T(ab)$  is not a simple arc it cannot be a simple closed curve. This follows either from (2.2) or from what was just shown. Hence  $T(ab)$  must con-

(7) See footnote 4.



tain a triod  $oc + od + oe = t$ . But, since  $T$  is cyclic chain-preserving,  $T(C(a, b))$  must contain a simple arc through the three points  $c, d, e$ , say  $cde$ . Then either  $cd$  or  $de$  does not contain  $o$ , say  $de$  does not contain  $o$ . Thus  $od + oe + de$  contains a simple closed curve  $J_b$  containing nondegenerate subarcs  $od'$  and  $oe'$  of  $od$  and  $oe$ ; and if  $E_b$  is the cyclic element of  $B$  containing  $J_b$ ,  $E_b \cdot T(ab) \supset E_b \cdot t \supset d'o' = od' + oe'$  and  $o$  is a limit point of  $T(ab) - d'e'$  contrary to what was shown above.

**3. Supplementary results.** We give here some additional results, throwing light on the action of arc-preserving transformations. We assume throughout this section that  $A$  is a compact locally connected continuum and that  $T(A) = B$  is arc-preserving.

(3.1) *The image of each  $A$ -set in  $A$  is an  $A$ -set in  $B$ ; the image of each cyclic chain in  $A$  is a cyclic chain in  $B$ .*

(3.2) *For each true cyclic element  $E_b$  of  $B$  there exists a unique true cyclic element  $E_a$  of  $A$  which maps onto  $E_b$  topologically under  $T$ .*

These are direct consequences of (2.6) and (2.7).

(3.3) *The image of every true cyclic element  $E_a$  of  $A$  is either a single point, a true cyclic element  $E_b$  of  $B$ , or a free arc of  $B$  which is also a cyclic chain of  $B$ . If  $T(E_a) = E_b$ , then  $T$  is topological on  $E_a$ .*

**Proof.** We have at once that  $T(E_a)$  is a cyclic chain  $C(x, y)$  in  $B$ , if we assume that it is not a single point. If  $C(x, y)$  is a single true cyclic element  $E_b$  of  $B$ , then our conclusion follows at once. Hence we may assume that  $x$  and  $y$  are distinct and that there exists at least one point  $z$  which separates  $x$  and  $y$  in  $B$ .

Let  $x'$  and  $y'$  be points of  $E_a$  mapping into  $x$  and  $y$ , respectively, and let  $J$  be a simple closed curve in  $E_a$  containing  $x'$  and  $y'$ , and define  $J' = T(J)$ . We first show that  $J'$  is a free arc of  $B$ . Regarding  $J$  as our space, we see that  $T$  is arc-preserving on  $J$ . If  $J'$  contains a true cyclic element  $F$  of itself, then by (3.2),  $T(J) = F$  is a homeomorphism. This is impossible since  $x$  and  $y$  are separated in  $B$  by the point  $z$ . Thus  $J'$  is a dendrite. But  $J'$  contains no triod, since if it did we could easily find an arc of  $J$  having a triod in its image. Therefore,  $J'$  is a simple arc.

Assume that  $J' = a'd'b'$  is not a free arc of  $B$ . Then there exists a triod  $t$  in  $B$  having  $d'$  as center and  $a', b'$  as two of its end points, where  $d'$  is some interior point of  $J'$ . Let  $c'$  be the other end point of  $t$ . Then the three arcs  $a'd', b'd', c'd'$  of  $t$  are disjoint except for  $d'$  and we may let  $\{c'_i\}$  be a sequence of distinct points on  $c'd'$  converging to  $d'$  as a limit. It follows at once that there exists a point  $d$  in  $T^{-1}(d')$  and a sequence of points  $\{c_i\}$  in  $\{T^{-1}(c'_i)\}$  converging to  $d$ .

Since  $A$  is a locally connected continuum, there exists a region  $R$  in  $A$

containing  $b$  but disjoint from both  $T^{-1}(a')$  and  $T^{-1}(b')$ . This region is arc-wise connected and hence contains a simple arc  $cd$ , where  $c$  is one of the points  $c_i$ .

If  $d$  lies on  $J$  we let  $z$  be the first intersection of  $cd$  with  $J$ , and consider an arc  $G$  defined as the sum of  $cz$  and an arc of  $J$  intersecting both  $T^{-1}(a')$  and  $T^{-1}(b')$ . It is immediate that  $T(G)$  contains a triod, which is impossible since  $T$  is arc-preserving. Thus  $d$  does not lie on  $J$ .

Let  $z$  be the first intersection of the arc  $cd$  with the closed set  $T^{-1}(J')$ , and define  $cxe$  as a simple arc in  $A$  having the unique point  $e$  in common with  $J$ , and on  $cxe$  let  $y$  be the last intersection with  $cz$ . Define an arc  $H$  as follows: (a') if  $y$  is not  $z$  then  $H$  is the sum of the subarc  $zy$  of  $zc$  and the subarc  $ye$  of  $cxe$ ; (b') if  $y$  is  $z$  then  $H$  is the sum of  $cz$  and the subarc  $ze$  of  $cxe$ . Let  $G$  be a simple arc of  $A$  composed of  $H$  and a subarc of  $J$  intersecting both  $T^{-1}(a')$  and  $T^{-1}(b')$ . Then  $T(G)$  contains a triod, which is impossible.

Thus  $J'$  is a free arc  $a'b'$  of  $B$ . Thus every point of the open subarc  $xy$  of  $a'b'$  must separate  $x$  and  $y$  in  $B$ . Accordingly,  $C(x, y) = xy$  and  $C(x, y)$  is a free arc of  $B$ .

(3.4)<sup>(8)</sup> If  $a$  and  $b$  are two points of  $A$  having the same image point under  $T$ , then no true cyclic element of the chain  $C(a, b)$  can map topologically under  $T$ . Thus each true cyclic element in  $C(a, b)$  maps into either a single point or a free arc of  $B$ .

**Proof.** Let  $T(a) = T(b)$  and suppose there is a cyclic element  $E_a$  of  $C(a, b)$  which maps onto a cyclic element  $E_b$  of  $B$  topologically. Let  $aqrb$  be a simple arc in  $A$  where  $aq \cdot E_a = q$ ,  $rb \cdot E_a = r$ . Then either  $aq$  or  $rb$  is nondegenerate, and we may suppose  $aq$  is nondegenerate ( $rb$  may or may not be nondegenerate). Then  $T(aq + rb) = K$  is a continuum. Furthermore,  $K \cdot E_b$  contains the two distinct points  $T(q)$  and  $T(r)$ . Hence  $K \cdot E_b$  contains a point  $x$  distinct from both  $T(q)$  and  $T(r)$  which is the image of an internal point  $x_0$  of  $E_a$ . This is impossible, since  $T^{-1}(x)$  also intersects  $aq + rb - (q + r)$ , whereas by (2.7) and (2.6),  $T^{-1}(x)$  must consist of a single point.

4. **Dendrite-preserving property of arc-preserving transformations.** If  $A$  is a dendrite (or tree) and  $T(A) = B$  is arc-preserving, then it follows at once from (2.2) that  $T$  is dendrite-preserving. This fact can also be seen from (2.5), since if  $B$  had a true cyclic element  $E_b$  then  $A$  would also have one. This dendrite-preserving property of arc-preserving transformations was first noted and proven by R. G. Simond<sup>(9)</sup> as mentioned in the introduction of this paper. However, it is interesting to note that it follows directly from (2.4) of A.P.T. by the reasoning just given above, since the irreducibility of  $T$

<sup>(8)</sup> This is closely related to (2.3) of A.P.T. and, indeed, yields (2.3) of A.P.T. as a special case.

<sup>(9)</sup> See footnote 3.

assumed in (2.4) of A.P.T. does not limit the generality because we can take a sub-dendrite of  $A$  on which  $T$  is irreducible.

In view of the considerable length and difficulty of Miss Simond's proof, the following one which is self-contained and independent of all other results on arc-preserving transformations may be of interest.

**THEOREM (Simond).** *Arc-preserving transformations are dendrite-preserving.*

**Proof.** Let  $T(A) = B$  be arc-preserving, where  $A$  is a dendrite. We first show:

I. (Whether  $A$  is a dendrite or not.) If  $t = oa + ob + oc$  is a triod such that  $T(o) = o'$ ,  $T(a) = a'$ ,  $T(b) = b'$ ,  $T(c) = c'$ , and if  $t \cdot T^{-1}(a' + b' + c') = a + b + c$ , then  $T(oa) \cdot T(ob) = T(oa) \cdot T(oc) = T(ob) \cdot T(oc) = T(o) = o'$ .

For if, say,  $T(oa) \cdot T(ob)$  contains a point  $q'$  distinct from  $o'$ , we may suppose the order  $a', o', q', b'$  on the arc  $a'o'b' = T(aob)$ . Then  $T(ao)$  is a subarc  $a'o'q'$  of  $a'o'b'$ . Hence the arc  $T(aoc)$  consists of  $a'o'q'$  plus an arc  $q'c'$  from  $q'$  to  $c'$  which contains neither  $o'$  nor  $b'$ . But then the arc  $T(boc) = b'o'c'$  would contain both the arc  $o'q'b'$  of  $a'o'b'$  and the arc  $q'c'$ , which is impossible since clearly  $o'q'b' + q'c'$  contains a triod. This proves I.

Now suppose, contrary to the theorem, that  $B$  has a true cyclic element  $B'$ . Let  $A'$  be a minimal  $A$ -set in  $A$  such that  $T(A')$  contains  $B'$ . Then since  $A'$  is a dendrite but not an arc, there exists a point  $o$  in  $A'$  and three continua  $X, Y, Z$  such that  $A' = X + Y + Z$  and  $X \cdot Y = Y \cdot Z = Z \cdot X = o$ . Let  $T(o) = o'$ . Since  $B'$  is cyclic and  $T(Y + Z)$  does not contain  $B'$ , there exists a point  $q'$  in  $B' - o'$  and points  $x$  in  $X, y$  in  $(Y + Z)$ , such that  $T(x) = T(y) = q'$ . Clearly we may suppose  $y$  in  $Y$ . Take the arcs  $xo$  and  $yo$  in  $X$  and  $Y$  respectively. Then since both  $T(xo)$  and  $Y(yo)$  contain arcs from  $o'$  to  $q'$  whereas  $T(xo + oy)$  must be a simple arc, clearly  $T(xo) \cdot T(oy)$  contains an arc from  $o'$  to  $q'$ . Hence there is no loss of generality in assuming (as we shall do) that both  $x$  and  $y$  are cut points of  $A'$ . Let  $R_x$  and  $R_y$  be components of  $A' - x$  and  $A' - y$  lying in  $X - x$  and  $Y - y$  respectively. Then since no one of the sets  $T(A' - R_x), T(A' - R_y), T(X + Y)$  contains  $B'$ , there exist points  $a', b',$  and  $c'$  in  $B'$  such that  $T^{-1}(a') \cdot A' \subset R_x, T^{-1}(b') \cdot A' \subset R_y, T^{-1}(c') \cdot A' \subset Z - o$ . Let  $a \in T^{-1}(a'), b \in T^{-1}(b'),$  and  $c \in T^{-1}(c')$  be so chosen that for the arcs  $ax, yb,$  and  $oc$  in  $A'$  we have  $ax \cdot T^{-1}(a') = a, yb \cdot T^{-1}(b') = b,$  and  $T^{-1}(c') \cdot oc = c$ . Let  $oa = ox + xa, ob = oy + yb$ . Then  $t = oa + ob + oc$  is a triod satisfying the conditions in I. However, since each of the sets  $T(oa)$  and  $T(ob)$  contains both  $o'$  and  $q'$  we have a contradiction to I. Thus  $B$  can have no true cyclic element and hence must be a dendrite.

5.  **$A$ -set reversing transformations.** In conclusion, we give a characterization of  $A$ -set reversing transformations which is made possible by (2.6) and (2.7). We recall that  $T(A) = B$  is  $A$ -set reversing provided that for each  $b$

in  $B$ ,  $T^{-1}(b)$  is either a single point or an  $A$ -set in  $A$ , it being assumed that  $A$  is a compact locally connected continuum. We make use here of certain results concerning this type of transformation which were established in §4 of A.P.T.

**THEOREM.** *If  $A$  is a compact locally connected continuum and  $T(A) = B$  is arc-preserving, then in order that  $T$  be  $A$ -set reversing it is necessary and sufficient that the following conditions hold: (a) there exists no true cyclic element  $E$  in  $A$  such that  $T(E)$  is a free arc of  $B$ ; (b) if  $K$  is the set of all cut points and end points of  $A$ , then  $T$  is monotone on  $T(K)$ .*

**Proof.** The necessity follows at once from the definition and the fact that every  $A$ -set reversing transformation is monotone (A.P.T., (4.12)).

**Sufficiency:** By (A.P.T., (4.1)) we must show that  $T$  is monotone on each simple arc in  $A$ . If this is not the case, there exists a simple arc  $axb$  in  $A$  such that  $T(a) = T(b) \neq T(x)$  for any point  $x$  interior to  $axb$ . If  $a$  and  $b$  are conjugate points, they lie in the same true cyclic element  $E$  of  $A$  and it is immediate from (a) and (3.3) that  $T$  is monotone on  $axb$ . Thus there exists a point  $q$  interior to  $axb$  which separates  $a$  and  $b$  in  $A$ . It follows at once from (b) that  $a$  is an internal point of a true cyclic element  $E$  of  $A$  which maps topologically onto a true cyclic element  $F$  of  $B$ . But  $T^{-1}$  is single valued on  $T(E^0)$ , where  $E^0$  represents the set of all internal points of  $E$ .

BROWN UNIVERSITY,  
PROVIDENCE, R. I.,  
THE UNIVERSITY OF VIRGINIA,  
CHARLOTTESVILLE, VA.

# ORTHOGONAL POLYNOMIALS WITH AUXILIARY CONDITIONS

BY  
DUNHAM JACKSON

1. **Introduction.** Let  $U_1(f)$ ,  $U_2(f)$ ,  $\dots$ ,  $U_m(f)$  be  $m$  linear functionals,  $m \geq 1$ , each defined for a class of functions  $f$  including all polynomials in a single variable  $x$ . The characterization of a functional  $U(f)$  as *linear* means here merely that if  $f_1$  and  $f_2$  are any two functions to which the operation applies,  $U(c_1f_1 + c_2f_2) = c_1U(f_1) + c_2U(f_2)$ . This paper is concerned with sets of polynomials  $p_n(x)$  orthogonal on an interval  $(a, b)$ , and satisfying the auxiliary conditions  $U_i(p_n) = 0$  for  $i = 1, 2, \dots, m$ , and for each value of  $n$ . Two of the simplest special cases, one corresponding to the single condition  $p_n(1) = p_n(-1)$  and the other to the condition  $p_n(1) = -p_n(-1)$ , have already been discussed elsewhere<sup>(1)</sup>. It will be shown here that certain formal propositions with regard to such orthogonal systems can be stated with a considerable degree of generality, while the theory of convergence is carried appreciably beyond the stage previously attained.

## 2. Construction of the orthogonal system. If

$$p_n(x) = a_{n0} + a_{n1}x + a_{n2}x^2 + \dots + a_{nn}x^n,$$

it follows from the property of linearity that

$$U_i(p_n) = \sum_{k=0}^n \gamma_{ik} a_{nk}, \quad \gamma_{ik} = U_i(x^k).$$

To the given set of auxiliary conditions there corresponds a matrix

$$(1) \quad \begin{array}{cccc} \gamma_{10} & \gamma_{11} & \gamma_{12} & \dots \\ \gamma_{20} & \gamma_{21} & \gamma_{22} & \dots \\ . & . & . & \dots \\ \gamma_{m0} & \gamma_{m1} & \gamma_{m2} & \dots \end{array}$$

Presented to the Society, September 7, 1939; received by the editors January 11, 1940.

<sup>(1)</sup> See D. Jackson, *A new class of orthogonal polynomials*, American Mathematical Monthly, vol. 46 (1939), pp. 493-497.

Since the present paper was written and since publication of the article in the Monthly, I have received through the kindness of Professor Mauro Picone a reprint of a paper by Wolfgang Gröbner, *Sistemi di polinomi ortogonali soddisfacenti a date condizioni*, number 62 of the Pubblicazioni dell'Istituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche, Rome, 1939, which also initiates a theory of orthogonal polynomials with linear homogeneous auxiliary conditions. That treatment and the one given here, however, diverge almost from the beginning as to methods and results to such an extent that there is very little duplication.



with  $m$  rows and infinitely many columns. Conversely, every such matrix, not consisting entirely of zeros, can be regarded as defining a set of  $m$  (not necessarily independent) linear homogeneous conditions  $U_i(p_n) = 0$ , significant for an arbitrary polynomial.

Let  $r_n$  be the rank of the matrix of the first  $n+1$  columns of (1), and let  $r_{-1} = 0$ . If  $r_n = r_{n-1}$ , there exist polynomials

$$P_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n,$$

with  $a_n \neq 0$ , satisfying the  $m$  conditions  $U_i(P_n) = 0$ . For if  $a_n$  is taken equal to 1, the relations to be satisfied by  $a_0, \dots, a_{n-1}$  are

$$\sum_{k=0}^{n-1} \gamma_{ik} a_k = -\gamma_{in}, \quad i = 1, 2, \dots, m,$$

and the condition  $r_n = r_{n-1}$  is precisely the condition that the matrix of this system of equations have the same rank as the augmented matrix. If  $r_n \neq r_{n-1}$ , that is, if  $r_n = r_{n-1} + 1$ , the equations are incompatible; the same is of course true if instead of 1 any other value different from zero is assigned to  $a_n$ , and there exists no polynomial of the  $n$ th degree with  $a_n \neq 0$  satisfying the conditions.

As  $n$  takes on the values  $0, 1, 2, \dots$ , since  $r_n$  can never decrease, can never increase by more than one unit at a time, and can never exceed  $m$ , there will be at most  $m$  values of  $n$  for which there is no polynomial satisfying the conditions. If  $r_n$  never attains the value  $m$ , the  $m$  conditions are linearly dependent; in the case of  $m$  independent conditions there are<sup>(2)</sup> just  $m$  exceptional values of  $n$ . It will be assumed henceforth that the conditions are independent.

Let polynomials satisfying the auxiliary conditions be constructed successively for all possible values of  $n$ , and let Schmidt's process be applied to these polynomials. It will be understood that the definition of orthogonality and normalization involves a weight function  $\rho(x)$  which is non-negative, and positive on a set of positive measure on  $(a, b)$ . Let the orthogonal polynomials when normalized be denoted by  $p_n(x)$ , the subscript indicating the degree of the polynomial in each case. For convenience of notation, let  $p_n(x) \equiv 0$  for the excluded values of  $n$ , and also for such negative values of  $n$  as may enter into any of the subsequent formulas.

Any polynomial of the  $n$ th degree satisfying the auxiliary conditions can be expressed linearly in terms of  $p_0, p_1, \dots, p_n$ . For terms in  $x^n, x^{n-1}, \dots$  can be removed successively by subtraction of multiples of  $p_n, p_{n-1}, \dots$ , leaving each time a polynomial which satisfies the conditions; when a degree is reached for which non-trivial polynomials satisfying the conditions do not

<sup>(2)</sup> See also Gröbner, loc. cit., p. 30, where the conclusion is stated with reference to a less general system of auxiliary conditions.



exist, the leading coefficient in the corresponding remainder must already be zero.

**3. Recursion formula and Christoffel-Darboux identity.** The ordinary procedure for setting up a recursion formula does not apply without modification, for if a polynomial satisfying the auxiliary conditions is multiplied by  $x$  the product does not satisfy the conditions in general. However, if each of the functionals  $U_i(f)$  is expressible in terms of the values of  $f$  at a finite number of points in the form

$$(2) \quad U_i(f) = C_{i1}f(y_1) + C_{i2}f(y_2) + \cdots + C_{ir}f(y_r),$$

where the  $y$ 's are real, or else conjugate complex in pairs with corresponding conjugate complex coefficients, the conditions  $U_i(f) = 0$  are satisfied by any polynomial which vanishes at  $y_1, y_2, \dots, y_r$ . (As a matter of notation, the list  $y_1, \dots, y_r$  is understood to include all the points that occur in any of the  $U$ 's; some of the coefficients  $C_{ij}$  may be zero.) If  $q(x)$  is the product

$$(x - y_1)(x - y_2) \cdots (x - y_r),$$

or any polynomial divisible by this product (or, with trivial increase of generality but with a possible slight gain in simplicity or convenience, a constant plus any such polynomial<sup>(\*)</sup>),  $q(x)p_n(x)$  satisfies the conditions for each value of  $n$ , and is expressible linearly in terms of  $p_0, \dots, p_{n+\mu}$ , where  $\mu \geq r$  is the degree of  $q(x)$ . By the property of orthogonality, the coefficient of  $p_k(x)$  in this representation is zero for  $k < n - \mu$ , and the representation has the form

$$q(x)p_n(x) = \sum_{k=n-\mu}^{n+\mu} c_{nk}p_k(x),$$

with

$$c_{nk} = \int_a^b \rho(x)q(x)p_n(x)p_k(x)dx.$$

These formulas hold for all non-negative integral values of  $n$  without exception, on the basis of the convention introduced above according to which  $p_k(x)$  is identically zero when not defined otherwise.

From the recursion formula a Christoffel-Darboux identity can be derived in the usual way.

Similar reasoning is possible if  $U_i(f)$  involves a finite number of derivatives at the points  $y_j$ . If  $e_j$  is the order of the highest derivative involved at  $y_j$ ,  $q(x)$  as defined above is to be replaced by

$$\prod_{j=1}^r (x - y_j)^{e_j+1},$$

(\*) E.g. in the earlier paper referred to, American Mathematical Monthly, loc. cit., the  $y$ 's being the points  $\pm 1$ ,  $x^2$  was used as multiplier instead of  $x^2 - 1$ .

or by a polynomial divisible by this product, or by a constant plus such a polynomial.

On the other hand, if there is just one auxiliary condition  $U_1(f)=0$ , where

$$U_1(f) = \int_{-1}^1 f(x) dx,$$

there is certainly no polynomial  $q(x)$  (other than a constant) such that  $q(x)p_n(x)$  satisfies the condition for all values of  $n$ . For that would require that  $q(x)$  be orthogonal to every polynomial whose integral over  $(-1, 1)$  is zero, and so orthogonal to every Legendre polynomial of positive degree, and such a polynomial is a constant. There is no recursion formula which expresses  $q(x)p_n(x)$  linearly in terms of the  $p$ 's for all  $n$ , with a polynomial factor  $q(x)$ .

In the case of a single auxiliary condition  $U_1(f)=0$ , with

$$U_1(f) = C_1 f(y_1) + C_2 f(y_2) + \cdots + C_r f(y_r),$$

the one exceptional value of  $n$  for which  $p_n(x) \equiv 0$ , the smallest value of  $n$  for which  $U_1(x^n) \neq 0$ , cannot exceed  $r-1$ . For the equations  $U_1(x^k)=0$ ,  $k=0, 1, \dots, r-1$ , constitute a set of linear equations for the  $C$ 's, having for its determinant the nonvanishing Vandermonde determinant of the powers of the  $y$ 's. That is to say,  $U_1(x^k)$  cannot vanish for all these values of  $k$  unless the  $C$ 's are all zero.

If there are  $m$  (linearly independent) conditions of the form (2), at least one  $m$ -rowed determinant of the first  $r$  columns of (1) must be different from zero;  $r_n = m$  for  $n \geq r-1$ , and  $p_n(x)$  is non-trivial for all values of  $n \geq r$ . For if all the  $m$ -rowed determinants in the first  $r$  columns were zero, the  $m$  sets of quantities  $U_i(x^k)$ ,  $k=0, 1, \dots, r-1$ , would be linearly dependent; there would be numbers  $b_1, \dots, b_m$ , not all zero, such that

$$\sum_{i=1}^m b_i (C_{i1} y_1^k + C_{i2} y_2^k + \cdots + C_{ir} y_r^k) = 0, \quad k = 0, 1, \dots, r-1;$$

that is,

$$C'_1 y_1^k + C'_2 y_2^k + \cdots + C'_r y_r^k = 0, \quad C'_i = \sum_{i=1}^m b_i C_{ii}.$$

By the argument of the preceding paragraph all the coefficients  $C'_i$  must vanish, which means that the  $m$  sets of coefficients  $C_{i1}, \dots, C_{ir}$  are linearly dependent.

**4. Boundedness of the normalized polynomials; convergence.** If  $f(x)$  is an integrable function on  $(a, b)$ , it can be formally expanded in a series of the polynomials  $p_n(x)$ , the coefficients being determined in the usual way. When there is a Christoffel-Darboux identity, it can be used for the study of con-

vergence in the same way as in connection with other orthogonal systems<sup>(4)</sup>, if the polynomials  $p_n(x)$  are bounded as  $n$  becomes infinite, at the point where convergence is to be proved.

The discussion of boundedness here will be not so much a general theory as an exploration of the effectiveness of particular types of hypothesis leading to the property in question. The auxiliary conditions will in each case involve the values of the polynomials, or of the polynomials and their derivatives, at only a finite number of points, and even with this limitation will be rather highly specialized in form. The interval of orthogonality will for simplicity be taken as  $(-1, 1)$ . The weight function, while open to subsequent generalization, will in the first instance be taken as unity.

Consider first the single condition  $p_n(y_1) = p_n(-y_1)$ . With  $y_1 = 1$ , this was treated in the earlier paper to which reference has been made. The condition is satisfied by any even polynomial, and by any polynomial which is divisible by  $x^2 - y_1^2$ . For  $n$  even let  $p_n(x)$  denote the normalized Legendre polynomial of the  $n$ th degree. There is no polynomial of the first degree satisfying the auxiliary condition. For odd  $n \geq 3$  let  $p_n(x) = (x^2 - y_1^2)\pi_{n-2}(x)$ , where  $\pi_k(x)$  denotes the polynomial of the  $k$ th degree in the orthonormal system corresponding to weight function  $(x^2 - y_1^2)^2$ . For  $n$  odd,  $\pi_{n-2}(x)$  is an odd polynomial, since the weight function is even. Inasmuch as any odd polynomial is orthogonal to any even polynomial, the even and odd  $p$ 's together constitute the desired orthogonal system. The normalized Legendre polynomials are uniformly bounded in any closed interval interior to  $(-1, 1)$ . The same is true<sup>(5)</sup> of the polynomials  $(x^2 - y_1^2)^2\pi_k(x)$ . Hence  $p_n(x)$  is similarly bounded for odd as well as even  $n$ , if  $y_1$  is not interior to the interval  $(-1, 1)$ , and is uniformly bounded in the interval except near the points  $\pm 1$ ,  $\pm y_1$ , if  $y_1$  is between  $-1$  and  $1$ .

Suppose there are two conditions,  $p_n(y_1) = p_n(-y_1)$ ,  $p_n(y_2) = p_n(-y_2)$ . They are satisfied by any even polynomial, and by any polynomial divisible by  $(x^2 - y_1^2)(x^2 - y_2^2)$ . They are not satisfied by any polynomial of the first or third degree. The orthogonal system consists of the normalized even Legendre polynomials and the polynomials  $(x^2 - y_1^2)(x^2 - y_2^2)\pi_{n-4}(x)$  with  $n$  odd, where  $\pi_k(x)$  denotes the general polynomial in the orthonormal system for weight  $(x^2 - y_1^2)^2(x^2 - y_2^2)^2$ . They are uniformly bounded throughout any closed interval interior to  $(-1, 1)$  and not containing any of the points  $\pm y_1$ ,  $\pm y_2$ . The extension to an arbitrary number of conditions of the form  $p_n(y_i) = p_n(-y_i)$  is obvious.

A set of conditions of the form  $p_n(y_i) = -p_n(-y_i)$  leads to similar results. For a single condition  $p_n(y_1) = -p_n(-y_1)$ , the orthonormal system consists of

<sup>(4)</sup> See e.g. D. Jackson, *Series of orthogonal polynomials*, Annals of Mathematics, (2), vol. 34 (1933), pp. 527-545; *Orthogonal trigonometric sums*, the same Annals, vol. 34 (1933), pp. 799-814; *A class of orthogonal functions on plane curves*, the same Annals, vol. 40 (1939), pp. 521-532.

<sup>(5)</sup> See e.g. D. Jackson, *Series of orthogonal polynomials*, loc. cit., pp. 534-535.

the normalized odd Legendre polynomials and the polynomials  $(x^2 - y_1^2)\pi_{n-2}(x)$ ,  $n = 2, 4, \dots$ , where  $\pi_0, \pi_2, \dots$  are the even orthonormal polynomials for  $(x^2 - y_1^2)^2$  as weight function.

With conditions of the form last mentioned, the ordinary proof of convergence, after the polynomials are known to be bounded, requires modification in one particular, because of the fact that the orthogonal system does not include a constant. Consider for definiteness the case of the single condition  $p_n(y_1) = -p_n(-y_1)$ . If  $f(x)$  is a function developed in series of the  $p$ 's, the partial sum of the series is given by

$$s_n(x) = \int_{-1}^1 f(t) K_n(t, x) dt, \quad K_n(t, x) = \sum_{k=1}^n p_k(t) p_k(x).$$

A polynomial of the  $n$ th or lower degree satisfying the auxiliary condition is reproduced by this formula exactly. For example,

$$x = \int_{-1}^1 t K_n(t, x) dt.$$

If  $f(x)$  can be represented in the form  $x\phi(x)$ , where  $\phi(x)$  is a function of sufficient regularity, convergence can be treated by means of the formulas

$$f(x) = x\phi(x) = \int_{-1}^1 \phi(t) t K_n(t, x) dt,$$

$$s_n(x) - f(x) = \int_{-1}^1 [\phi(t) - \phi(x)] t K_n(t, x) dt.$$

The assumption that  $f(x)$  can be represented in the form  $x\phi(x)$  is no essential restriction, as far as convergence at other points than  $x=0$  is concerned, for it can be seen as in other cases that convergence at a point depends only on the behavior of the function in the neighborhood of the point.

Occasion arises for a somewhat less simple treatment of the problem in connection, for example, with the auxiliary condition  $p_n'(1) = p_n'(-1)$ . This is satisfied by any odd polynomial; it is not satisfied by any polynomial of the second degree, but it is satisfied by a constant, or by any polynomial divisible by  $(1-x^2)^2$ . The even polynomials of the orthogonal system, however, do not consist merely of a constant and the polynomials  $(1-x^2)^2 q_k(x)$ , where the  $q$ 's are orthogonal for weight  $(1-x^2)^4$ ; for example,  $(1-x^2)^2 q_0(x)$  is not orthogonal to a constant.

Let  $p_0(x), p_1(x), p_2(x), p_3(x), \dots$  be the orthonormal polynomials satisfying the auxiliary condition, and let  $\xi_0(x), \xi_1(x), \xi_2(x), \dots$  be the normalized Legendre polynomials. The odd  $p$ 's are the odd  $\xi$ 's. (It is readily seen, as in other problems having analogous features of symmetry, that the  $p$ 's of even degree are even polynomials, and those of odd degree are odd.) For  $n$  even let

$$(3) \quad p_n(x) = c_{n0}\xi_0(x) + c_{n1}\xi_1(x) + \cdots + c_{nn}\xi_n(x),$$

$$c_{nk} = \int_{-1}^1 p_n(x)\xi_k(x)dx.$$

The polynomial  $\xi_k(x) - \frac{1}{2}x^2\xi_k'(1)$ , with  $k$  even, has a vanishing derivative for  $x = \pm 1$ , and satisfies the auxiliary condition. Hence  $p_n(x)$  is orthogonal to it when  $n > k$ :

$$\int_{-1}^1 p_n(x)\xi_k(x)dx - \xi_k'(1) \int_{-1}^1 \frac{1}{2}x^2 p_n(x)dx = 0, \quad k = 0, 2, \dots, n-2,$$

i.e., if the last integral is denoted by  $g_n$ ,  $c_{nk} = g_n \xi_k'(1)$ .

If  $P_k(x)$  is the non-normalized Legendre polynomial, so that  $\xi_k(x) = [(2k+1)/2]^{1/2}P_k(x)$ ,

$$P_k'(1) = k(k+1)/2, \quad \xi_k'(1) = k(k+1)(2k+1)^{1/2}/2^{3/2}.$$

Since  $p_n(x)$  is normalized,

$$1 = \int_{-1}^1 [p_n(x)]^2 dx = \sum_{k=0}^n c_{nk}^2 = c_{nn}^2 + \frac{1}{2}g_n^2 \sum_{k=0}^{n-2} k^2(k+1)^2(2k+1),$$

the sign  $\sum'$  indicating summation over even values of  $k$ . The sum by which  $g_n^2$  is multiplied is of the order of magnitude of  $n^6$ , from which it follows that  $g_n = O(1/n^3)$ .

Let

$$\alpha_k = [2/(2k+1)]^{1/2}\xi_k'(1) = k(k+1)/2,$$

$$S_n(x) = \sum_{k=0}^{n-2} \xi_k'(1)\xi_k(x) = \sum_{k=0}^{n-2} \alpha_k [(2k+1)/2]^{1/2}\xi_k(x) = \sum_{k=0}^{n-2} \alpha_k \xi_k(1)\xi_k(x).$$

Let

$$\sigma_k(x) = \sum_{j=0}^k \xi_j(1)\xi_j(x)$$

for even  $k$ . Then ( $\alpha_0$  being zero)

$$S_n(x) = \sum_{k=2}^{n-2} \alpha_k [\sigma_k(x) - \sigma_{k-2}(x)] = - \sum_{k=0}^{n-4} (\alpha_{k+2} - \alpha_k)\sigma_k(x) + \alpha_{n-2}\sigma_{n-2}(x).$$

Now, with summation extended over both odd and even values of  $k$ ,

$$\sum_{j=0}^k \xi_j(1)\xi_j(x) = \frac{k+1}{[(2k+1)(2k+3)]^{1/2}} \frac{\xi_{k+1}(1)\xi_k(x) - \xi_k(1)\xi_{k+1}(x)}{1-x},$$

which, as  $\xi_k(1) = O(k^{1/2})$ ,  $\xi_k(x) = O(1)$ , does not exceed a constant multiple



of  $k^{1/2}$  on a closed interval interior to  $(-1, 1)$ . A similar statement holds if  $x$  is replaced by  $-x$ , and consequently holds for the even and odd parts of the sum separately. In particular,  $|\sigma_k(x)| = O(k^{1/2})$  uniformly in any closed interval interior to  $(-1, 1)$ . On the other hand,  $\alpha_{k+2} - \alpha_k = O(k)$ . So  $|S_n(x)| = O(n^{5/2})$ .

The relation (3) may be written

$$p_n(x) = c_{nn}\xi_n(x) + g_n S_n(x).$$

From the preceding paragraphs,  $|g_n S_n(x)| = O(1/n^{1/2})$ . By application of Schwarz's inequality to the integral defining the coefficient,  $|c_{nn}| \leq 1$ . The polynomials  $p_n(x)$  are uniformly bounded over any closed interval interior to  $(-1, 1)$ .

An essentially similar problem is that associated with the condition  $p'_n(1) = -p'_n(-1)$ . The polynomials of even degree in the orthogonal system are Legendre polynomials. The requisite information about the coefficients in the representation of the odd polynomials of the system in terms of Legendre polynomials comes from the fact that  $\xi_k(x) - x\xi'_k(1)$  satisfies the auxiliary condition when  $k$  is odd.

Considerations of the same sort are effective in connection with the unsymmetric condition  $p_n(1) = h p_n(-1)$ , where  $h$  is an arbitrary constant  $\neq \pm 1$ . Here the orthogonal polynomials are neither even nor odd<sup>(\*)</sup>. In the representation

$$p_n(x) = \sum_{k=0}^n c_{nk} \xi_k(x)$$

the coefficients  $c_{nk}$  for  $k < n$  are determined in accordance with the fact that  $\xi_k(x) - \mu \xi_k(1)$  satisfies the auxiliary condition with  $\mu = (1+h)/(1-h)$  when  $k$  is odd, and  $\xi_k(x) - \xi_k(1)$  satisfies it when  $k$  is even. Hence, for  $k < n$ ,  $c_{nk} = g_n \xi_k(1)$  or  $\mu g_n \xi_k(1)$  according as  $k$  is even or odd, with

$$g_n = \int_{-1}^1 p_n(x) dx.$$

Since

$$\sum_{k=0}^n c_{nk}^2 = 1, \quad \xi_k(1) = [(2k+1)/2]^{1/2},$$

it follows that  $g_n = O(1/n)$ . And as was noted above,  $|\sum \xi_k(1) \xi_k(x)| = O(n^{1/2})$  in the interior of  $(-1, 1)$ , whether the summation is extended over all subscripts from 0 to  $n$ , or over the even subscripts or the odd subscripts of the set separately. Hence the desired conclusion with regard to the boundedness of the  $p$ 's.

The condition  $p_n(1) = 0$  leads merely to the set of polynomials

(\*) For an explicit determination of these polynomials see Gröbner, loc. cit., pp. 46-47.



$(x-1)\pi_{n-1}(x)$ , where the  $\pi$ 's are orthonormal for weight  $(x-1)^2$ . The condition  $p'_n(1)=0$  appears to be less trivial; boundedness of the  $p$ 's can be proved by use of the observation that  $\xi_k(x)-x\xi'_k(1)$  has a vanishing derivative for  $x=1$ .

A primitive example of higher order is the condition  $p''_n(1)=0$ . It is satisfied by  $\xi_k(x)-\frac{1}{2}x^2\xi''_k(1)$ . So  $p_n(x)$  is orthogonal to this expression for  $k < n$ , and if  $p_n(x) = \sum_k c_{nk}\xi_k(x)$ , then for  $k < n$

$$c_{nk} = g_n \xi''_k(1), \quad g_n = \int_{-1}^1 \frac{1}{2} x^2 p_n(x) dx.$$

Since  $\xi''_k(1) = \frac{1}{8}(k+2)(k+1)k(k-1)\xi_k(1)$ , it follows by reasoning similar to that which has already been used that  $g_n = O(1/n^5)$ ,  $|\sum \xi''_k(1)\xi_k(x)| = O(n^{9/2})$  in the interior of the interval, and the  $p$ 's are bounded as in other cases.

Consider next the pair of conditions  $p_n(1)=p_n(-1)$ ,  $p'_n(1)=p'_n(-1)$ . Because of the symmetry of the problem, the orthogonal polynomials are even or odd, and the even and odd sequences can be considered separately. When  $n$  is even, the conditions are satisfied by  $\xi_n(x)-\frac{1}{2}x^2\xi'_n(1)$ ; when  $n$  is odd they are satisfied by  $\xi_n(x)-x\xi_n(1)$ . In each case it follows on the basis of calculations which have been presented already that the  $p$ 's are bounded except near the ends of the interval.

As a final illustration, of somewhat more general character, suppose there is a single auxiliary condition  $U_1(p_n)=0$  expressed in terms of the values of  $p_n$  and an arbitrary finite number of its derivatives at the points  $\pm 1$ . Let  $x^\lambda$  be a power of  $x$ , for simplicity the lowest power, such that  $U_1(x^\lambda) \neq 0$ . Let

$$\psi(x) = \xi_k(x) - A_k x^\lambda.$$

The coefficient  $A_k$  can be determined so that

$$U_1(\psi) = U_1[\xi_k(x)] - A_k U_1(x^\lambda) = 0.$$

When  $A_k$  is thus determined,  $p_n(x)$  is orthogonal to  $\psi(x)$  if  $n > k$ ,  $n > \lambda$ :

$$\int_{-1}^1 p_n(x) \xi_k(x) dx = A_k g_n, \quad g_n = \int_{-1}^1 x^\lambda p_n(x) dx.$$

Since  $\xi_k^{(h)}(-1) = (-1)^{k+h} \xi_k^{(h)}(1)$ , there is a set of coefficients  $a_0, a_1, \dots, a_\alpha$ , independent of  $k$ , with  $a_\alpha \neq 0$ , such that

$$A_k = a_0 \xi_k(1) + a_1 \xi'_k(1) + \dots + a_\alpha \xi_k^{(\alpha)}(1)$$

when  $k$  is even, unless  $A_k=0$  for all even values of  $k$ , and a set of coefficients  $b_0, b_1, \dots, b_\beta$ , independent of  $k$ , with  $b_\beta \neq 0$ , such that

$$A_k = b_0 \xi_k(1) + b_1 \xi'_k(1) + \dots + b_\beta \xi_k^{(\beta)}(1)$$

when  $k$  is odd, unless  $A_k = 0$  for all odd values of  $k$ ; here  $\alpha$  and  $\beta$  are in general equal to the order  $\gamma$  of the highest derivative occurring in  $U_1(p_n)$ , but one of them may in particular have a smaller value. At least one of the numbers  $\alpha, \beta$  is certainly equal to  $\gamma$ , and at least one coefficient  $a_\gamma$  or  $b_\gamma$  is present with a value different from zero.

The quantity  $\xi_k^{(\gamma)}(1)$  is of the order of magnitude of  $k^{2\gamma+(1/2)}$ , and  $\sum_{k=0}^{n-1} A_k^2$  is not less than a positive constant multiple of  $n^{4\gamma+2}$  when  $n$  is sufficiently large. From the fact that  $\sum_k c_{nk}^2 = 1$ , if  $p_n(x) = \sum_k c_{nk} \xi_k(x)$ , it follows that  $g_n = O(1/n^{2\gamma+1})$ . Since

$$[\xi_{k+1}^{(\gamma)}(1)/\xi_{k+1}(1)] - [\xi_k^{(\gamma)}(1)/\xi_k(1)] = O(k^{2\gamma-1}),$$

it may be shown by the use of partial summation in conjunction with the Christoffel identity, in the manner previously indicated, together with inequalities obtained in the same way or, more simply, without resort to partial summation for the derivatives of lower order, that

$$\left| \sum_{k=0}^{n-1} A_k \xi_k(x) \right| = O(n^{2\gamma+(1/2)})$$

in the interior of  $(-1, 1)$ . The  $p$ 's are bounded as before.

It is apparent that the methods that have been used are capable of further extension. It is not so clear what the most general explicit formulation would be. On the other hand, it may be that some different method would lead to more general results at a single stroke.

THE UNIVERSITY OF MINNESOTA,  
MINNEAPOLIS, MINN.

# CONTINUOUS ADDITIVE FUNCTIONALS ON THE SPACE $(BV)$ AND CERTAIN SUBSPACES

BY

C. RAYMOND ADAMS AND ANTHONY P. MORSE

1. **Introduction.** We consider here the class  $(BV)$  of functions  $x(t)$  of bounded variation on the interval

$$I = E[0 \leq t \leq 1].$$

Its intersection with the class  $(C)$  of functions continuous on  $I$  will be designated by  $(CBV)$ , its subclass of absolutely continuous functions by  $(AC)$ , and  $(BV) - (CBV)$  by  $(DBV)$ . In a recent paper<sup>(1)</sup> Adams introduced for  $(BV)$  the metric

$$(1) \quad (x, y) = \int_0^1 |x(t) - y(t)| dt + |T_0^1(x) - T_0^1(y)|,$$

$T_0^1(z)$  being employed in general to denote the total variation of the function  $z(t)$  on  $I$ . Thus metrised,  $(BV)$  is not a Banach space<sup>(2)</sup>; but it is complete, separable, and boundedly compact. Although a linear space, it is not a "linear topological space" in the sense in which that term is sometimes used, for the topology introduced by the metric (1) is non-uniform. Indeed it is easily seen that the category of a subset is not always invariant under translation. For, if the closed unit sphere  $K(\theta, 1)$  about the zero-element  $\theta$  as center, which is a set of second category in  $(BV)$ , is subjected to the translation  $x$  where  $x$  has as a representative function  $x(t) = 0$  for  $0 < t \leq 1$ ,  $x(0) = 2$ , then its translate

$$E = E[y = x + z, z \in K]$$

is a subset of  $(DBV)$  and so of first category<sup>(3)</sup> in  $(BV)$ . Regarded as a group,  $(BV)$  is discontinuous.

Presented to the Society, December 2, 1939; received by the editors January 25, 1940.

<sup>(1)</sup> Adams, *The space of functions of bounded variation and certain general spaces*, these Transactions, vol. 40 (1936), pp. 421-438, hereinafter referred to as A. The properties of  $(BV)$  mentioned presently are either explicitly established in, or easily to be inferred from the results of, this paper and its sequel by Adams and Morse, *On the space  $(BV)$* , *ibid.*, vol. 42 (1937), pp. 194-205, later referred to as AM.

<sup>(2)</sup> Indeed it is clear, from the fact that "convergence in variation" is not additive, that it is impossible to norm the set  $(BV)$ , or either of its subsets  $(CBV)$  and  $(AC)$ , in such manner that convergence in the metric determined by the norm is equivalent to convergence in the metric (1). See Adams and Clarkson, *On convergence in variation*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 413-417. This same remark holds for  $(BV)$  or  $(CBV)$  metrised with the distance function (17); see §6 below.

<sup>(3)</sup> See AM, p. 199. An example of a residual set which under the same translation goes into a set of first category is provided by  $(CBV)$ .

2. **Functionals on  $(BV)$ .** The functional<sup>(4)</sup>  $f(x) = \text{ess} \lim_{t \rightarrow 0} x(t) = \lim_{t \rightarrow 0} x(t)$  clearly is defined for every  $x \in (BV)$  in the natural sense that if  $x(t)$  is an element of the class  $(BV)$ ,  $f(x)$  is a real number, and if  $x(t)$  and  $y(t)$  are both elements of the class  $(BV)$ ,  $(x, y) = 0$  implies  $f(x) = f(y)$ . This functional is additive and homogeneous on  $(BV)$ , and it may readily be seen to be continuous at each point of the subset  $(BVN)$  corresponding to functions having no external saltus anywhere<sup>(5)</sup>; nevertheless it is discontinuous at each point of  $(BV) - (BVN)$ . Incidentally, both  $(BVN)$  and its complement are dense in  $(BV)$ . In further contrast to the situation in the case of a Banach space, there exist functionals which are additive and continuous on  $(BV)$  without being uniformly continuous. But any functional  $f(x)$  which is additive and uniformly continuous on  $(BV)$  does satisfy a Lipschitz condition,  $|f(x) - f(y)| \leq M \cdot (x, y)$  for  $x, y \in (BV)$ ; and the smallest number  $M$  which can be used in this inequality we have called the "modulus" of  $f$  on  $(BV)$  and designated by the symbol  $\text{mod}_{(BV)} f$ .

In A, Theorems 5.1, 5.2, and 5.3, it was shown that every functional  $f$  additive and uniformly continuous on  $(BV)$  [or on  $(CBV)$  or on  $(AC)$ ] can be expressed in the form of a Lebesgue integral,

$$(2) \quad \int_0^1 x(t) \alpha(t) dt, \quad \text{ess} \sup_{t \in I} |\alpha(t)| = M < \infty,$$

with  $\text{mod}_{(BV)} f = M$  [or  $\text{mod}_{(CBV)} f = M$  or  $\text{mod}_{(AC)} f = M$ ]; and that each integral of this kind is such a functional<sup>(6)</sup>. An example of an additive and continuous functional which is not uniformly continuous on  $(BV)$  is provided by any such integral with  $\alpha(t)$  summable but not essentially bounded. The general form of the additive and continuous, but not necessarily uniformly continuous, functional on  $(BV)$ , however, was not determined in A. This open question we now propose to settle.

**THEOREM 1.** *The conditions  $T_0^1(x_n) < B < \infty$  ( $n=0, 1, 2, \dots$ ),  $\lim_{n \rightarrow \infty} \int_0^1 |x_n - x_0| dt = 0$ , and  $g \in (C)$  imply  $\lim_{n \rightarrow \infty} \int_0^1 x_n dg = \int_0^1 x_0 dg$ .*

**Proof.** This theorem is equivalent to Bray's extension<sup>(7)</sup> of a theorem of

<sup>(4)</sup> By *functional* we mean an operation or transformation whose range is contained in the real number system. We recall the well known fact that in a Banach space continuity of an additive functional  $f$  at one point alone implies continuity everywhere, uniform continuity, and the satisfaction by  $f$  of a Lipschitz condition on the entire space.

<sup>(5)</sup> For a precise definition of  $(BVN)$  see the first paragraph of §3 below.

<sup>(6)</sup> More recently Hildebrandt, in *Linear operations on functions of bounded variation*, Bulletin of the American Mathematical Society, vol. 44 (1938), p. 75, has determined the general form of the continuous additive functional on the non-separable Banach space which the class  $(BV)$  becomes when normed with  $\|x\| = |x(0)| + T_0^1(x)$ . As in the case of other non-separable Banach spaces previously considered by this author, the functional is expressed by a generalized integral of Stieltjes or Lebesgue type which he constructs for the purpose.

<sup>(7)</sup> See Bray, *Elementary properties of the Stieltjes integral*, Annals of Mathematics, (2), vol. 20 (1918-1919), p. 180.

Helly, in the sense that each can be derived from the other. It seems preferable to us, however, to prove our result di novo rather than to use Bray's theorem as a basis. That the first two conditions imply uniform boundedness of  $x_n$ , and that all three imply  $\lim_{n \rightarrow \infty} \int_0^1 x_n dg = \int_0^1 x_0 dg$  in the particular case in which  $g \in (AC)$ , has already been remarked in the first paragraph of the proof of Theorem 5.1 of A. We now extend the proof to the general case, in which  $g$  is an arbitrary continuous function. Let  $\epsilon$  be any positive number; let  $B_1 \geq B$  be a bound for  $|x_n(t) - x_0(t)|$  ( $t \in I$ ;  $n = 1, 2, 3, \dots$ ); and let  $h(t)$  satisfy the conditions

$$h \in (AC), \quad \sup_{t \in I} |g(t) - h(t)| \leq \epsilon/(2B_1), \quad g(0) - h(0) = g(1) - h(1) = 0.$$

In accordance with the particular case of the theorem already proved we have  $\lim_{n \rightarrow \infty} \int_0^1 (x_n - x_0) dh = 0$ , whence as  $n \rightarrow \infty$

$$\begin{aligned} \limsup \left| \int_0^1 (x_n - x_0) dg \right| &\leq \limsup \left| \int_0^1 (x_n - x_0) d(g - h) \right| \\ &\quad + \limsup \left| \int_0^1 (x_n - x_0) dh \right| \\ &= \limsup \left| \int_0^1 (g - h) d(x_n - x_0) \right| \\ &\leq \limsup \frac{\epsilon}{2B_1} T_0^1(x_n - x_0) \\ &\leq \limsup \frac{\epsilon}{2B_1} [T_0^1(x_n) + T_0^1(x_0)] \leq \epsilon. \end{aligned}$$

COROLLARY. *Each integral*

$$(3) \quad \int_0^1 x(t) dg(t) \quad \text{with } g \in (C)$$

*is a continuous additive functional for  $x \in (BV)$ .*

THEOREM 2. *Each continuous additive functional on  $(BV)$  can be expressed in the form (3).*

Proof. Let  $f$  be any such functional and, as in the proof of Theorem 5.2 of A, set

$$\xi_t(u) = \begin{cases} 1 & \text{for } 0 \leq u \leq t, \\ 0 & \text{for } t < u \leq 1, \end{cases} \quad f(\xi_t) = g(t), \quad t \in I.$$

For any pair of numbers  $t, t_1$  in the interval  $0 \leq t < 1$  we have  $(\xi_t, \xi_{t_1}) = \int_0^1 |\xi_t - \xi_{t_1}| du$ , so that  $t \rightarrow t_1$  implies  $(\xi_t, \xi_{t_1}) \rightarrow 0$ . This in turn implies

$f(\xi_i) \rightarrow f(\xi_1)$ , since  $f$  is continuous at  $\xi_1$ ; i.e.,  $t \rightarrow t_1$  implies  $g(t) \rightarrow g(t_1)$ . If we let  $\eta(u) = 1$  for  $0 \leq u < 1$ ,  $\eta(1) = 0$ , it is clear that  $t \rightarrow 1$  implies  $(\xi_t, \eta) = \int_0^1 |\xi_t - \eta| du \rightarrow 0$ , which implies  $f(\xi_t) \rightarrow f(\eta)$ ; i.e., we may infer that  $\lim_{t \rightarrow 1} g(t)$  exists. Hence, defining  $\bar{g}(t) = g(t)$  for  $0 \leq t < 1$ ,  $\bar{g}(1) = \lim_{t \rightarrow 1} g(t)$ , we have  $\bar{g} \in (C)$  and  $\int_0^1 x d\bar{g}$  exists for every  $x \in (BV)$ . The argument used in the proof of Theorem 5.2 of A now holds, with no change whatever, from the beginning of the second paragraph up to and including equation (5.4); that is to say, it may be concluded that  $g(1) = \bar{g}(1)$  and  $f(x) = \int_0^1 x(t) dg(t)$  for every  $x \in (BV)$ .

3. **Functionals on  $(CBV)$  and  $(AC)$ .** An arbitrary function  $x$  will be said to have *no external saltus* if and only if at each point  $t_1 \in I$ ,  $x$  satisfies the condition  $\liminf_{t \rightarrow t_1} x(t) \leq x(t_1) \leq \limsup_{t \rightarrow t_1} x(t)$ . We shall employ  $(BVN)$  to designate the intersection of the class  $(BV)$  with the class of functions having no external saltus. Clearly  $x \in (BVN)$  implies continuity of  $x$  at  $t = 0$  and  $t = 1$ .

**THEOREM 3.** *The conditions<sup>(\*)</sup>  $x_n \in (BV)$  ( $n = 1, 2, 3, \dots$ ),  $x_0 \in (BVN)$ ,  $\lim_{n \rightarrow \infty} (x_n, x_0) = 0$ ,  $|g(t)| < B < \infty$  for  $t \in I$ , and  $\int_0^1 g dx_n$  exists<sup>(\*)</sup> ( $n = 0, 1, 2, \dots$ ) imply  $\lim_{n \rightarrow \infty} \int_0^1 g dx_n = \int_0^1 g dx_0$  and  $\lim_{n \rightarrow \infty} \int_0^1 x_n dg = \int_0^1 x_0 dg$ .*

**Proof.** In the same manner in which it may be seen that a closed set  $E \subset I$  can be inclosed in a *finite* set of disjoint intervals  $O_i$  each open with respect to  $I$  and the sum of whose lengths exceeds the measure of  $E$  by arbitrarily little, one may see that  $E$  can be inclosed in a *finite* set of such intervals with  $\sum_i T_{\bar{O}_i}(x_0)$  exceeding the variation<sup>(\*)</sup> of  $x_0$  on  $E$  by an arbitrarily small amount.

Let  $\epsilon$  be an arbitrary positive number,  $k$  a positive number satisfying the inequality  $kT_{\bar{O}_i}^1(x_0) < \epsilon$ , and  $D_k \subset I$  the set of points where  $g$  has a saltus  $\geq k$ . Since  $D_k$  is closed and the variation of  $x_0$  on  $D_k$  is zero,  $D_k$  can be inclosed in a finite set of disjoint intervals  $O_i$ , each open with respect to  $I$ , such that  $\sum_i T_{\bar{O}_i}(x_0) < \epsilon/(2B)$ . Since the points of continuity of  $x_0$  are dense in  $I$  and include 0 and 1, and since  $D_k$  is closed, each interval  $O_i$  can be shrunk (if necessary) into an interval  $O'_i$ , open with respect to  $I$ , whose end-points are points of continuity of  $x_0$  and such that  $D_k$  is still inclosed in  $\sum_i O'_i$ . By

(\*) From the proof it will be clear that a weaker set of conditions sufficient to insure the conclusion is obtained by replacing  $(x_n, x_0) \rightarrow 0$  by the following:  $\int_0^1 |x_n - x_0| dt \rightarrow 0$  and the existence of a set  $D$  dense in  $I$  and containing 0 and 1 and of a non-decreasing function  $F$  such that  $\int_0^1 g dF$  exists and on every closed interval  $J \subset I$  with end-points in  $D$ ,  $\limsup_{n \rightarrow \infty} T_J(x_n)$  does not exceed the increment of  $F$  on  $J$ . We take occasion to remark that this theorem neither includes nor is included by a theorem of Daniell on passage to the limit, *Further properties of the general integral*, *Annals of Mathematics*, (2), vol. 21 (1919-1920), p. 218.

(\*) It is desirable to recall here the meaning of the term "variation of a function on a set." Let  $x \in (BV)$ , and  $E \subset I$ ; then the variation of  $x$  on  $E$  is by definition the infimum of numbers of the form  $\sum_i T_{\bar{O}_i}(x)$  where  $E \subset \sum_i O_i$ , each  $O_i$  is an interval open with respect to  $I$  (i.e.,  $O_i$  is the intersection with  $I$  of some open interval), and  $\bar{O}_i$  is the closure of  $O_i$ . If  $g$  is bounded on  $I$  and  $x \in (BV)$ , a necessary and sufficient condition for  $\int_0^1 g dx$  to exist is that the variation of  $x$  on the set of points of discontinuity of  $g$  be zero. See, for example, Hobson, *Theory of Functions of a Real Variable*, 3d edition, vol. 1, Cambridge, 1927, p. 542.



Theorem 1 of AM,  $(x_n, x_0) \rightarrow 0$  implies the same condition on each subinterval  $\bar{O}'_i$ . Letting  $\alpha = \sum_i \bar{O}'_i$ ,  $T_\alpha(x_n) = \sum_i T_{\bar{O}'_i}(x_n)$  ( $n=0, 1, 2, \dots$ ), we therefore have as  $n \rightarrow \infty$

$$\begin{aligned} \limsup \left| \int_\alpha g dx_n - \int_\alpha g dx_0 \right| &\leq \limsup \left| \int_\alpha g dx_n \right| + \limsup \left| \int_\alpha g dx_0 \right| \\ &\leq \limsup B \cdot [T_\alpha(x_n) + T_\alpha(x_0)] \\ &= 2BT_\alpha(x_0) < \epsilon. \end{aligned}$$

Let  $\beta$  denote the finite set of disjoint closed intervals constituting the closure of the point set  $I - \alpha$ , at no point of which  $g$  has a saltus  $\geq k$ . By aid of the Heine-Borel theorem<sup>(10)</sup> one may easily see that on each interval of  $\beta$ ,  $g$  can be approximated uniformly within  $k/2$  by a continuous function  $h$ . We then have as  $n \rightarrow \infty$

$$\begin{aligned} \limsup \left| \int_\beta g dx_n - \int_\beta g dx_0 \right| &= \limsup \left| \int_\beta (g - h) dx_n + \int_\beta h dx_n - \int_\beta (g - h) dx_0 - \int_\beta h dx_0 \right| \\ &\leq \limsup \left| \int_\beta (g - h) dx_n - \int_\beta (g - h) dx_0 \right| + \limsup \left| \int_\beta h dx_n - \int_\beta h dx_0 \right| \\ &\leq \limsup (k/2) [T_\beta(x_n) + T_\beta(x_0)] = kT_\beta(x_0) < \epsilon, \end{aligned}$$

since  $\lim_{n \rightarrow \infty} \int_\beta h dx_n = \int_\beta h dx_0$  is an immediate consequence of Theorem 1, the formula for integration by parts, and the fact that by Theorem 2 of AM we have<sup>(11)</sup> pointwise convergence of  $x_n(t)$  to  $x_0(t)$  at each end-point of the intervals constituting  $\beta$ . That  $\int_0^1 x_n dg \rightarrow \int_0^1 x_0 dg$  now follows at once by aid of the formula for integration by parts and the pointwise convergence of  $x_n(t)$  to  $x_0(t)$  at  $t=0$  and  $t=1$ .

<sup>(10)</sup> An explicit construction for  $h$ , on an interval which may as well be taken as  $I$ , is the following. Let  $M(t)$ ,  $N(t)$  be the "maximum and minimum functions" for  $g$  (i.e., for example, let  $M(t) = \lim_{\delta \rightarrow 0} \sup_{t, |t_1 - t| < \delta} g(t_1)$  for  $t \in I$ ); the saltus of  $g$  at  $t$  is then  $M(t) - N(t)$ , and this is  $< k$  for  $t \in I$ . Setting  $M_n(t) = \sup_{t_1 \in I} [f(t_1) - n|t_1 - t|]$ ,  $N_n(t) = \inf_{t_1 \in I} [f(t_1) + n|t_1 - t|]$  ( $t \in I$ ;  $n = 1, 2, 3, \dots$ ), we easily see that for each  $n$  these functions satisfy a Lipschitz condition, and that  $M_n(t) \rightarrow M(t)$  from above and  $N_n(t) \rightarrow N(t)$  from below as  $n \rightarrow \infty$ , so that  $M_n(t) - N_n(t)$  tends to  $M(t) - N(t)$  from above. Let

$$E_n = E_t [M_n(t) - N_n(t) \geq k];$$

then each  $E_n$  is closed, each  $E_n \supset E_{n+1}$ , and  $\prod_{n=1}^\infty E_n$  is vacuous. Hence there exists an integer  $n_0$  for which  $E_{n_0}$  is vacuous; i.e., we have  $M_{n_0}(t) - N_{n_0}(t) < k$  for  $t \in I$ . The desired function  $h$  may now be taken to be  $[M_{n_0}(t) + N_{n_0}(t)]/2$  for  $t \in I$ .

<sup>(11)</sup> According to this theorem the conditions  $x_n \in (BV)$  ( $n = 1, 2, 3, \dots$ ),  $x_0 \in (BVN)$ , and  $(x_n, x_0) \rightarrow 0$  imply pointwise convergence of  $x_n$  to  $x_0$  at each point of continuity of  $x_0$ .

For convenience in stating the following corollaries we let  $(R)$  stand for the class of functions which are Riemann integrable on  $I$  and  $(R^*)$  for the subclass of  $(R)$  of which each function has only a *countable* number of discontinuities. Recalling<sup>(12)</sup> that each pair of conditions,  $x \in (CBV)$ ,  $g \in (R^*)$  and  $x \in (AC)$ ,  $g \in (R)$ , is sufficient to insure the existence of  $\int_0^1 x dg$ , we have

COROLLARY 1. *Each integral*

$$(4) \quad \int_0^1 x(t) dg(t) \quad \text{with } g \in (R^*)$$

is a continuous additive functional for  $x \in (CBV)$ .

COROLLARY 2. *Each integral*

$$(5) \quad \int_0^1 x(t) dg(t) \quad \text{with } g \in (R)$$

is a continuous additive functional for  $x \in (AC)$ .

In conjunction with Theorem 2, Corollary 1 shows that there exist continuous additive functionals on  $(CBV)$  which cannot be extended to be continuous and additive on  $(BV)$ . In determining the general form of such a functional on  $(CBV)$  it is therefore desirable, if not actually necessary, to work wholly within the space  $(CBV)$  itself. We propose to prove that the general form of such a functional on  $(CBV)$  is (4) and on  $(AC)$  is (5). For this purpose we shall employ two lemmas as follows. Allowing  $(AC)_0$  to denote the subset of  $(AC)$  of which each function vanishes at  $t=0$ , we have

LEMMA 1. *Let  $f$  be a functional whose domain includes  $(AC)_0$ . If  $f$  is additive on  $(AC)_0$  and continuous on  $(AC)_0$ , metrised with the distance function  $(x, y) = T_0^1(x-y)$ , there exists a function  $h$  bounded and summable on  $I$  such that*

$$(6) \quad f(x) = \int_0^1 h(t) x'(t) dt \quad \text{for } x \in (AC)_0.$$

This result is an immediate consequence of the facts that  $(AC)_0$  is isometric with the Banach space  $(L)$  of functions summable on  $I$  and that for  $x' \in (L)$  the general form of the continuous additive functional  $f(x')$  on  $(L)$  is given<sup>(13)</sup> by the integral in (6). Since convergence in the metric  $(x, y) = T_0^1(x-y)$  implies convergence in the metric (1), any functional  $f$  whose domain includes  $(AC)_0$  and which is additive and continuous on  $(AC)_0$  metrised with (1) can be expressed, for  $x \in (AC)_0$ , in the form (6).

To promote our later convenience we formulate the

<sup>(12)</sup> See, for example, Hobson, loc. cit., p. 545.

<sup>(13)</sup> See, for example, Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 65.

DEFINITION. The function  $h$  in (6) shall be called the normalized function associated with the functional  $f$  if and only if it has the properties

$$\begin{aligned} h(0) &= f(x_1) \text{ where } x_1(t) = 1 \text{ for } t \in I, & h(1) &= 0, \\ (7) \quad h(t) &= \frac{1}{2} \left[ \limsup_{\delta \rightarrow 0} \int_t^{t+\delta} h(u) du / \delta + \liminf_{\delta \rightarrow 0} \int_t^{t+\delta} h(u) du / \delta \right] \\ &\text{for } 0 < t < 1. \end{aligned}$$

It should be clear from Lemma 1 and from this definition that associated with each functional  $f$  of the kind specified there is a normalized function  $h$ ; and that if  $\epsilon$  is an arbitrary positive number and  $t_1$  an arbitrary point in the open interval  $0 < t < 1$ , there exists in every neighborhood of  $t_1$  a set of positive measure on which  $h(t)$  is  $< h(t_1) + \epsilon$  and a set of positive measure on which  $h(t)$  is  $> h(t_1) - \epsilon$ .

It follows at once that any functional  $f$  whose domain includes  $(AC)$ , and which is additive and continuous on  $(AC)$  metrised with the distance function  $(x, y) = |x(0) - y(0)| + T_0^1(x - y)$ , can be expressed in the form<sup>(14)</sup>

$$(8) \quad f(x) = x(0)f(x_1) + \int_0^1 h(t)x'(t)dt = x(0)h(0) + \int_0^1 h(t)x'(t)dt \text{ for } x \in (AC),$$

where  $h$  has the properties (7).

LEMMA 2. Let  $\delta$  be an arbitrary number  $> 0$ ;  $P$  a non-vacuous perfect set  $\subset I$ ;  $x(t)$  a continuous non-decreasing function<sup>(15)</sup> with  $x(0) = 0$ ,  $x(1) = 1$ , and  $x'(t) = 0$  for  $t \in I - P$ ; and  $h(t)$  a bounded measurable function for  $t \in I$ , with essential saltus  $\geq k > 0$  at each point of  $P$ . Then there exist two non-decreasing functions  $\lambda(t), \mu(t) \in (AC)_0$  satisfying the conditions

$$\begin{aligned} \sup_{t \in I} |\lambda(t) - x(t)| &< \delta, & \sup_{t \in I} |\mu(t) - x(t)| &< \delta, \\ (9) \quad \int_0^1 h(t)\lambda'(t)dt - \int_0^1 h(t)\mu'(t)dt &\geq k/2. \end{aligned}$$

<sup>(14)</sup> The usual scheme for determining the general form of the continuous additive functional  $f$  on the Banach space  $(C)$  employs the device of extending  $f$  to points outside of  $(C)$ ; i.e., to the class of step-functions or to the entire space  $(M)$  of essentially bounded measurable functions. It may therefore be of interest to note that we now have in hand a means of obtaining the form of  $f$  without going outside of  $(C)$  itself. In fact from (8), by following essentially the same procedure as is used below in the proof of Lemma 3 (only that in the present instance  $x_1$  should be taken as  $x_1(t) = 0$  for  $t \in I$ ,  $\epsilon$  should be taken as 1, and we have no concern with the values of  $\int_0^1 |x_{\delta, \epsilon}(t)| dt$  and  $L_0^1(x_{\delta, \epsilon})$ ), one may readily show  $\sup_{x \in (AC), \|x\| = 1} |f(x)| \geq T_0^1(h)$ . Letting  $g(t) = h(0) - h(t)$  for  $t \in I$ , and observing that  $(AC)$  is dense in  $(C)$ , we may conclude  $f(x) = \int_0^1 x dg$  for  $x \in (C)$  and  $\|f\|_{(C)} = T_0^1(g) = T_0^1(h)$ .

<sup>(15)</sup> One readily sees that, for any non-vacuous  $P$ , there exists a function  $x$  on  $I$  with these properties; for example, in any subinterval of  $I$  where  $P$  is dense,  $x(t)$  may be taken as  $t$ , and elsewhere it may be defined by essentially the same process as is used for defining the Cantor ternary function.

**Proof.** Let  $N$  be a positive integer large enough so that

$$(10) \quad |x(t_1) - x(t_2)| < \epsilon \quad \text{for } |t_1 - t_2| < 1/N,$$

and let  $I_n$  ( $n=1, 2, \dots, N$ ) stand for the interval  $(n-1)/N < t < n/N$ . For each  $n$  we define functions  $\lambda_n, \mu_n$  as follows. If  $P \cdot I_n$  is vacuous, set  $\lambda_n(t) = \mu_n(t) = 0$  for  $t \in I$ . If  $P \cdot I_n$  is non-vacuous, set

$$S_n = \text{ess sup}_{t \in I_n} h(t), \quad s_n = \text{ess inf}_{t \in I_n} h(t),$$

so that  $S_n - s_n \geq k$ ; let  $\alpha_n, \beta_n$  be any measurable subsets of  $I_n$  with

$$\begin{aligned} |\alpha_n| &> 0, & |S_n - h(t)| &< k/4 \quad \text{for } t \in \alpha_n, \\ |\beta_n| &> 0, & |s_n - h(t)| &< k/4 \quad \text{for } t \in \beta_n; \end{aligned}$$

and set

$$\lambda_n(t) = \begin{cases} \Delta_{I_n} x / |\alpha_n| & \text{for } t \in \alpha_n, \\ 0 & \text{for } t \in I - \alpha_n, \end{cases} \quad \mu_n(t) = \begin{cases} \Delta_{I_n} x / |\beta_n| & \text{for } t \in \beta_n, \\ 0 & \text{for } t \in I - \beta_n, \end{cases}$$

where the notation  $|E|$  stands for the measure of a set  $E$  and  $\Delta_{I_n} x$  represents the increment of the function  $x$  on the closure of the interval  $I_n$ . We observe

$$\begin{aligned} \int_0^1 h(t) \lambda_n(t) dt &= \int_{\alpha_n} h(t) \lambda_n(t) dt \\ (11) \quad &\geq \int_{\alpha_n} (S_n - k/4) \lambda_n(t) dt = (S_n - k/4) \Delta_{I_n} x, \\ \int_0^1 h(t) \mu_n(t) dt &\leq (s_n + k/4) \Delta_{I_n} x, \end{aligned}$$

and we assert that the non-decreasing functions

$$\lambda(t) = \int_0^t \sum_{n=1}^N \lambda_n(u) du, \quad \mu(t) = \int_0^t \sum_{n=1}^N \mu_n(u) du$$

have the properties (9). It is clear that  $\lambda(t) = \mu(t) = x(t)$  for  $t = n/N$  ( $n=0, 1, \dots, N$ ), whence in view of (10) the first two of inequalities (9) are satisfied. As for the third, we have by (11)

$$\begin{aligned} \int_0^1 h(t) [\lambda'(t) - \mu'(t)] dt &= \int_0^1 h(t) \sum_{n=1}^N [\lambda_n(t) - \mu_n(t)] dt \\ &= \sum_{n=1}^N \int_0^1 h(t) \lambda_n(t) dt - \sum_{n=1}^N \int_0^1 h(t) \mu_n(t) dt \\ &\geq \sum_{n=1}^N (S_n - k/4 - s_n - k/4) \Delta_{I_n} x \geq (k/2) \sum_{n=1}^N \Delta_{I_n} x = k/2. \end{aligned}$$

**THEOREM 4.** *Each continuous additive functional on (CBV) can be expressed in the form (4).*

**Proof.** Each functional  $f(x)$  of this kind is expressible, for  $x \in (AC)$ , in the form (8). The function  $h$  can have an essential discontinuity at no more than a countable set of points; for the contrary would imply the existence of a number  $k > 0$  such that the points where  $h$  has an essential saltus  $\geq k$  would be a non-countable closed set, this set would contain a non-vacuous perfect set, and by Lemma 2 there would exist points  $x \in (CBV)$  at which  $f$  fails to be continuous. Being normalized,  $h$  is continuous whenever it is essentially continuous; hence  $h$  is continuous except at a countable set of points and we may write

$$x(0)h(0) + \int_0^1 h(t)x'(t)dt = x(0)h(0) + \int_0^1 h(t)dx(t).$$

This expression may be brought into the form (4) by setting  $g(t) = h(0) - h(t)$ .

**THEOREM 5.** *Each continuous additive functional on (AC) can be expressed in the form (5).*

This result can be demonstrated in the same manner as Theorem 4, the conclusion that  $h$  cannot have essential discontinuities at a set  $D$  of measure  $> 0$  being drawn from Lemma 2. For, if  $|D| > 0$ , there would exist a  $k > 0$  such that the set  $D_k \subset D$  where  $h$  has an essential saltus  $\geq k$  would be closed and of measure  $> 0$ ; and  $D_k$  would contain a perfect set  $P$  with  $|P| > 0$ . The function  $x$  of Lemma 2 could then be taken as  $\int_0^t \phi(u)du/|P|$ , where  $\phi$  is the characteristic function of  $P \subset I$ ; i.e., there would exist points  $x \in (AC)_0$  at which  $f$  fails to be continuous.

**4. Norms of the functionals.** By Theorem 2.3 of A, any functional  $f$  additive and continuous on (BV) satisfies a Lipschitz condition at the zero-element  $\theta \in (BV)$ , with a Lipschitz modulus which was called the "norm" of  $f$  on (BV) and designated by the symbol  $\|f\|_{(BV)}$ . On the assumption that  $g(0) = 0$ , which can be made without loss of generality, the upper bounds

$$(12) \quad T_0^1(g), \quad |g(1)| + \sup_{t \in I} |g(t)|$$

and the lower bounds

$$(13) \quad \frac{1}{2} \operatorname{osc}_{t \in I} g(t), \quad |g(1)|, \quad \frac{1}{2} \sup_{t \in I} |g(t)|$$

for  $\|f\|_{(BV)}$  were determined<sup>(16)</sup> in A. These show that if  $g$  is monotone on  $I$ ,  $\|f\|_{(BV)} = T_0^1(g)$ ; in no other case, however, was the norm evaluated. We now propose to evaluate the norms of the functionals (3), (4), and (5).

<sup>(16)</sup> See A, pp. 437-438. It was tacitly assumed there that the functional  $f$  under consideration was *uniformly* continuous on (BV); but this assumption was not used, the bounds being determined solely from the Stieltjes integral form of the functional.

For convenience in this connection we adopt the following conventions. If  $g$  is a function on  $I$  and

$$J = E[t' \leq t \leq t''] \subset I,$$

we define  $\Delta_J' g = g(t'') - g(t')$  when each of the points  $t', t''$  is an end-point of  $I$  or a point of continuity of  $g$ ,  $\Delta_J' g = 0$  otherwise;  $|J|$  will stand for the length of  $J$ ; and  $\nu(J)$  will be 0, 1, or 2 according as both, one, or neither of the conditions  $0 \in J$ ,  $1 \in J$  is satisfied. We then have

THEOREM 6. The norm of each of the functionals (3), (4), and (5) is

$$(14) \quad \sup_J |\Delta_J' g| / [|J| + \nu(J)]$$

as  $J$  ranges over the set of closed subintervals of  $I$ .

**Proof.** Since  $(AC) \subset (CBV)$  is dense in  $(CBV)$  and therefore in  $(BV)$ , we see that proving the norm of the functional (5) to be given by (14) is tantamount to proving the theorem. We proceed to consider, then, the functional (5).

Let  $(S_\theta)$  represent the class of step-functions<sup>(17)</sup> each of which is continuous at each point of discontinuity of  $g$ . Since the points of continuity of  $g$  are dense in  $I$ , we infer that  $(S_\theta) \subset (BV)$  is dense relative to  $(AC)$ . Now  $\int_0^1 x dg$  exists for  $x \in (AC) + (S_\theta)$ , and we define

$$F(x) = \int_0^1 x dg \quad \text{for } x \in (AC) + (S_\theta).$$

Clearly we have  $F(x) = f(x)$  for  $x \in (AC)$ ; and since  $(S_\theta)$  is  $\subset (BVN)$ , we conclude from Theorem 3 that  $F$  is continuous on  $(AC) + (S_\theta)$ . Thus we obtain<sup>(18)</sup>

$$\begin{aligned} \|f\|_{(AC)} &= \sup_{x \in (AC), \|x\|=1} \int_0^1 x dg = \sup_{x \in (S_\theta), \|x\|=1} \int_0^1 x dg \\ &= \sup_{x \in (S_\theta), \|x\| \leq 1} \int_0^1 x dg = \sup_{x \in (S_\theta), \|x\| > 0} \int_0^1 x dg / \|x\|. \end{aligned}$$

Let  $\epsilon$  be any positive number and  $y \in (S_\theta)$ , with  $\|y\| = 1$ , satisfy the inequality

$$\int_0^1 y dg > \|f\|_{(AC)} - \epsilon;$$

$y$  shall now be regarded as fixed. Let  $(S'_\theta)$  be the set of step-functions  $x$  defined by the condition  $x \in (S'_\theta)$  if and only if  $x$  is continuous at each point  $t$  where  $y$

<sup>(17)</sup> It should be clearly understood that by a step-function we mean here a function consisting of a finite number of steps each of length  $> 0$ .

<sup>(18)</sup> See A, p. 430. We use the notation  $\|x\| = (x, \theta) = \int_0^1 |x(t)| dt + T_0^1(x)$ .



is continuous. Then  $\int_0^1 x dg$ , for  $x \in (S'_\theta)$  and  $\|x\| = 1$ , is a continuous bounded function of the heights of the steps in  $x$  and so assumes a maximum; thus there exists a particular step-function  $x_0 \in (S'_\theta)$ , with  $\|x_0\| = 1$ , such that

$$\int_0^1 x dg \leq \int_0^1 x_0 dg \quad \text{for } x \in (S'_\theta), \|x\| = 1,$$

which implies

$$\int_0^1 x dg / \|x\| \leq \int_0^1 x_0 dg / \|x_0\| \quad \text{for } x \in (S'_\theta), \|x\| > 0.$$

Let  $t_0$  satisfy the condition

$$|x_0(t)| \leq |x_0(t_0)| \quad \text{for } t \in I,$$

and set

$$t'_0 = \inf_t E[x_0(u) = x_0(t_0) \text{ for } t < u < t_0],$$

$$t''_0 = \sup_t E[x_0(u) = x_0(t_0) \text{ for } t_0 < u < t],$$

$$J_0 = E[t'_0 \leq t \leq t''_0].$$

Finally, let  $x_\lambda \in (S'_\theta)$  be defined thus

$$x_\lambda(t) = \begin{cases} (1 + \lambda)x_0(t_0) & \text{for } t \in J_0, \\ x_0(t) & \text{for } t \in I - J_0. \end{cases}$$

Then we have

$$\int_0^1 x_\lambda dg = \int_0^1 x_0 dg + \lambda \cdot x_0(t_0) \cdot \Delta'_{J_0} g;$$

and since  $|x_0(t)|$  is actually *greater* for  $t \in J_0$  than it is for  $t$  immediately to the left or right of  $J_0$ , there exists an  $\eta > 0$  such that we have

$$\|x_\lambda\| = \|x_0\| + \lambda \cdot |x_0(t_0)| \cdot |J_0| + \lambda \cdot |x_0(t_0)| \cdot \nu(J_0) \quad \text{for } \lambda > -\eta.$$

The function

$$H(\lambda) = \frac{\int_0^1 x_\lambda dg}{\|x_\lambda\|} = \frac{\int_0^1 x_0 dg + \lambda \cdot [x_0(t_0) \cdot \Delta'_{J_0} g]}{\|x_0\| + \lambda \cdot [|x_0(t_0)| \cdot |J_0| + |x_0(t_0)| \cdot \nu(J_0)]}$$

is of the form  $(a + b\lambda)/(c + d\lambda)$  with  $cd > 0$ , and it has a maximum at  $\lambda = 0$ . Hence we have  $H'(0) = (ad - bc)/c^2 = 0$  and  $a/c = b/d$ ; i.e.,

$$\int_0^1 x_0 dg / \|x_0\| = |\Delta'_{J_0} g| / [|J_0| + \nu(J_0)].$$

Combining our results we obtain

$$\|f\|_{(AC)} - \epsilon < \int_0^1 y \, dg \leq \int_0^1 x_0 \, dg = \int_0^1 x_0 \, dg / \|x_0\| = |\Delta'_0 g| / [|J_0| + \nu(J_0)];$$

and since  $\epsilon$  is an arbitrary positive number, we conclude that  $\|f\|_{(AC)}$  is not greater than the number (14).

On the other hand, if  $J$  is any closed subinterval of  $I$  and the function  $x_J$  is defined by

$$x_J(t) = \begin{cases} \operatorname{sgn}(\Delta'_J g) / [|J| + \nu(J)] & \text{for } t \in J, \\ 0 & \text{for } t \in I - J, \end{cases}$$

we see at once

$$x_J \in (S_g), \quad \|x_J\| \leq 1, \quad \int_0^1 x_J \, dg = |\Delta'_J g| / [|J| + \nu(J)],$$

whence

$$\begin{aligned} |\Delta'_J g| / [|J| + \nu(J)] &\leq \sup_{x \in (S_g), \|x\| \leq 1} \int_0^1 x \, dg \\ &= \sup_{x \in (S_g), \|x\| = 1} \int_0^1 x \, dg = \|f\|_{(AC)}. \end{aligned}$$

Thus  $\|f\|_{(AC)}$  is not less than the number (14), and the theorem is proved.

It may be worth while to point out here that the formula (14) provides a good basis for computation. For example, in the case of  $g(t) = 4t(1-t)$ , the value (14) is assumed for

$$J = J_0 = E[0 \leq t \leq 2^{1/2} - 1],$$

for which  $\nu(J_0) = 1$ , and the norm is  $4(2^{1/2} - 1)(2 - 2^{1/2}) / 2^{1/2} = .69$  approximately. In the case of  $g(0) = g(1) = 0$ ,  $g(.49) = -10$ ,  $g(.51) = 10$  and  $g$  linear on each of the closed intervals  $[0, .49]$ ,  $[.49, .51]$ ,  $[.51, 1]$ , the value (14) is assumed for

$$J = J_0 = E[.49 \leq t \leq .51],$$

for which  $\nu(J_0) = 2$ , and the norm is  $20 / 2.02 = 9.90$  approximately. Minor variants of this example show that  $\|f\|_{(BV)}$  can be arbitrarily close to  $|g(1)| + \sup_{t \in I} |g(t)|$ ; and other examples to indicate that each of the estimates (12) and (13) is in a sense the best possible can readily be constructed.

We may observe also that the inequality  $|T_0^1(x) - T_0^1(y)| \leq T_0^1(x - y)$  implies that if  $x_1$  is an arbitrary point in  $(BV)$ , we have

$$\sup_{x \in (BV), (x, x_1) > 0} |f(x) - f(x_1)| / (x, x_1) \\ \geq \sup_{x \in (BV), (x, x_1) > 0} |f(x - x_1)| / (x - x_1, \theta) \geq \|f\|_{(BV)};$$

i.e., that the Lipschitz modulus of a continuous additive functional  $f$  at any point in the space is never less than its Lipschitz modulus at the zero-element  $\theta$ .

5. **Weak topologies in  $(BV)$ .** The two forms of functionals (3) and (2) respectively provide the basis for the following

**DEFINITIONS.** A sequence  $x_n$  ( $n = 1, 2, 3, \dots$ ) of elements of  $(BV)$  will be said to converge weakly (S) [to converge weakly (W)] if and only if  $\lim_{n \rightarrow \infty} f(x_n)$  exists for every functional  $f$  additive and continuous [additive and uniformly continuous] on  $(BV)$ .

It is clear that convergence of a sequence  $x_n$  in the metric (1) implies that  $x_n$  converges weakly (S), and that convergence of  $x_n$  weakly (S) implies convergence of  $x_n$  weakly (W). That implications do not hold in the reverse direction may be seen from more or less trivial examples. From (2), which is also the general form of the continuous additive functional on the Banach space  $(L)$ , as has been remarked in §3, it is clear that the weak (W) topology of  $(BV)$  is equivalent to the topology introduced in  $(BV) \subset (L)$  by the weak topology of  $(L)$ . Since  $(BV) \subset (L)$  is strongly dense in  $(L)$ , it is apparent that the weak closure of  $(BV) \subset (L)$  is  $(L)$ . It has been shown earlier that  $(L)$  is weakly complete<sup>(19)</sup>.

**THEOREM 7.** In the topology of weak (S) convergence,  $(BV)$  is complete.

**Proof.** Let  $x_n$  ( $n = 1, 2, 3, \dots$ ) be any sequence in  $(BV)$  satisfying the condition

$$(15) \quad \lim_{n \rightarrow \infty} \int_0^1 x_n dg \text{ exists} \quad \text{for every } g \in (C);$$

and let  $\bar{x}_n$  be a function associated with  $x_n$  ( $n = 1, 2, 3, \dots$ ) as follows:

$$\bar{x}_n(t) = \begin{cases} \lim_{t_1 \rightarrow t, t_1 > t} x_n(t) & \text{for } 0 < t < 1, \\ 0 & \text{for } t = 0, t = 1. \end{cases}$$

Then we have  $\int_0^1 x_n dg = \int_0^1 \bar{x}_n dg$  for every  $g \in (C)$  and every  $n$ . For fixed  $n$ ,

$$(16) \quad \int_0^1 \bar{x}_n dg = - \int_0^1 g d\bar{x}_n = \int_0^1 g d(-\bar{x}_n) \quad (n = 1, 2, 3, \dots)$$

is a continuous additive functional<sup>(20)</sup> on the Banach space  $(C)$ , with norm equal to  $T_0^1(-\bar{x}_n) = T_0^1(\bar{x}_n)$ . The condition (15) implies that each  $g \in (C)$  has a

<sup>(19)</sup> See Banach, loc. cit., pp. 141-142.

<sup>(20)</sup> That is, linear functional, in the sense of Banach, loc. cit.

bounded sequence of images under the sequence of transformations (16). It follows from a theorem of Banach and Steinhaus<sup>(21)</sup> that the sequence of norms  $T_0^1(\bar{x}_n)$  ( $n=1, 2, 3, \dots$ ) is bounded. Therefore the sequence of functions  $\bar{x}_n$  is uniformly bounded, and from a theorem of Helly<sup>(22)</sup> we conclude the existence of a subsequence  $\bar{x}_{n_i}$  ( $i=1, 2, 3, \dots$ ) which converges pointwise for all  $t \in I$  to a function  $x_0 \in (BV)$ . From Lebesgue's convergence theorem we infer that  $\bar{x}_{n_i}$  converges in the mean to  $x_0$ ; and from Theorem 1 above we conclude that  $\int_0^1 \bar{x}_n dg$ , and therefore  $\int_0^1 x_n dg$ , tends to  $\int_0^1 x_0 dg$  for every  $g \in (C)$ .

One may readily verify the following remarks.

(i) In each of the weak topologies, the weak limit of a sequence  $x_n$  ( $n=1, 2, 3, \dots$ ) in  $(BV)$  is not unique in the sense of uniqueness determined by metric equality in  $(BV)$ ; it is, however, unique in the space  $(L)$ .

(ii) In contrast to the situation in a Banach space<sup>(23)</sup>, boundedness of the sequence  $\|x_n\|$  ( $n=1, 2, 3, \dots$ ), where  $\|x_n\| = (x_n, \theta)$ , is not a necessary condition for weak convergence, in either sense, of a sequence  $x_n$  in  $(BV)$ ; but boundedness of the sequence  $\int_0^1 |x_n(t)| dt$  ( $n=1, 2, 3, \dots$ ) is of course necessary.

#### 6. The use of the metric

$$(17) \quad (x, y) = \int_0^1 |x(t) - y(t)| dt + |L_0^1(x) - L_0^1(y)|,$$

where  $L_0^1(z)$  stands in general for the (Peano) length of the function  $z(t)$  on  $I$ .

When this metric is employed the situation is as described in the following

**THEOREM 8.** *Each uniformly continuous additive functional on  $(BV)$  [or on  $(CBV)$  or on  $(AC)$ ] can be expressed in the form (2). The general form of the continuous additive functional on  $(BV)$  is (3), on  $(CBV)$  is (4), and on  $(AC)$  is (8). Conversely, each integral of the kind specified is such a functional on the corresponding space. The functional (8) on  $(AC)$  satisfies a Lipschitz condition at any given point  $x_1 \in (AC)$  if and only if  $h(t)$  satisfies a Lipschitz condition on  $I$ ; in this event the integral (8) can be expressed in the form (2), with  $g(t) = h(0) - h(t)$ , and this integral defines a functional  $f$  on  $(BV)$  which satisfies a Lipschitz condition on the entire space  $(BV)$ ; and the Lipschitz modulus of  $f$  on  $(BV)$  [or on  $(CBV)$  or on  $(AC)$ ] is the same at each point  $x_1$  of  $(BV)$  [or of  $(CBV)$  or of  $(AC)$ ] as it is for the entire space, being equal to the Lipschitz modulus of  $g(t) = h(0) - h(t)$  on  $I$ .*

We shall endeavor to indicate the proof of these results without going into excessive detail. Naturally it must be noted at the outset that since the metric (17) is not homogeneous (i.e., does not satisfy the condition  $(ax, ay)$ )

<sup>(21)</sup> See, for example, Banach, loc. cit., p. 80.

<sup>(22)</sup> See Helly, *Über lineare Funktionaloperationen*, Sitzungsberichte der Wiener Akademie, IIa, vol. 121 (1912), p. 283.

<sup>(23)</sup> See, for example, Banach, loc. cit., p. 133.

$= |a| \cdot (x, y)$ , for  $a$  a real number), the spaces considered are neither of the type  $(\alpha)$  nor of the type  $(\alpha^*)$ , so that the results of §2 of A cannot be drawn upon. In particular, there would seem to be no a priori reason why a uniformly continuous additive functional should satisfy a Lipschitz condition<sup>(24)</sup>; that such is the case, however, will presently appear. It is a simple matter to show that a continuous additive functional on  $(BV)$  or one of its subspaces here considered is homogeneous<sup>(25)</sup>.

Let  $f$  be any additive and uniformly continuous functional on  $(BV)$ . As in the proof of Theorem 5.2 of A and of Theorem 2 above, introduce the family of step-functions  $\xi_t(u)$  and define  $f(\xi_t) = g(t)$ . In view of the uniform continuity of  $f$ , let  $\delta > 0$  be such that

$$|f(x) - f(y)| < 1 \quad \text{for } (x, y) \leq \delta.$$

Let  $0 \leq t_1 < t_2 < 1$  and let  $m$  satisfy the condition

$$m \int_0^1 |\xi_{t_1} - \xi_{t_2}| du = \int_0^1 |m\xi_{t_1} - m\xi_{t_2}| du = \delta.$$

Then we have

$$m |g(t_1) - g(t_2)| = m |f(\xi_{t_1}) - f(\xi_{t_2})| = |f(m\xi_{t_1}) - f(m\xi_{t_2})| < 1,$$

whence

$$|g(t_1) - g(t_2)| < 1/m = \int_0^1 |\xi_{t_1} - \xi_{t_2}| du / \delta = |t_1 - t_2| / \delta;$$

i.e.,  $g(t)$  satisfies a Lipschitz condition on the interval  $0 \leq t < 1$ . Consider

$$x_1(t) = \begin{cases} 0 & \text{for } 0 \leq t < 1, \\ k & \text{for } t = 1, \end{cases} \quad x_2(t) = \begin{cases} 0 & \text{for } 0 \leq t < 1, \\ -k & \text{for } t = 1. \end{cases}$$

Since  $(x_1, x_2) = 0$  and  $(x_1, -x_2) = 0$ , we have  $f(x_1) = f(x_2) = f(-x_2) = -f(x_2) = 0$ ; and from the additivity of  $f$  it follows that the value of  $f(x)$  is independent of the value of  $x(1)$ . The sequence of step-functions  $x_n(t)$  employed in the proof of Theorem 5.2 of A will then have the property that in the present metric  $(x_n, \bar{x}) \rightarrow 0$ , with  $\bar{x}$  identical with  $x$  except at  $t = 1$  where it differs from  $x$  by  $1 + T_0^1(x) - L_0^1(x)$ . The argument set forth in that proof then shows that  $f(x)$  can be expressed as  $\int_0^1 x(t) dg(t)$ , where  $g(t)$  is Lipschitzian on  $I$ ; i.e.,  $f(x)$  can be given the form (2), where  $\text{ess sup}_{t \in I} |\alpha(t)| = M$  is the Lipschitz modulus of  $g$  on  $I$ . The existence of a Lipschitz modulus for  $f$  on  $(BV)$ , and its

<sup>(24)</sup> For example, the linear functionals  $f(x) = kx$  ( $k \neq 0$ ) on the Euclidean space  $E_1$  metrised with  $(x, y) = |x^2 - y^2|$ , which is not homogeneous, do not satisfy a Lipschitz condition.

<sup>(25)</sup> Compare the reasoning used in the proof of Theorem 2.4 of A and make use of the relation  $|L_0^1(x) - L_0^1(y)| \leq T_0^1(x - y)$ ; see inequality (4) of Adams and Lewy, *On convergence in length*, Duke Mathematical Journal, vol. 1 (1935), pp. 19-26.

equality to  $M$ , then follows from the relation

$$\begin{aligned} M &= \sup_{0 \leq t_1 < t_2 \leq 1} |g(t_1) - g(t_2)| / |t_1 - t_2| \\ &= \sup_{0 \leq t_1 < t_2 < 1} |f(\xi_{t_1}) - f(\xi_{t_2})| / |t_1 - t_2| \\ &= \sup_{0 \leq t_1 < t_2 < 1} |f(\xi_{t_1}) - f(\xi_{t_2})| / (\xi_{t_1} - \xi_{t_2}) \\ &\leq \sup_{x, y \in (BV), (x, y) > 0} |f(x) - f(y)| / (x, y) \\ &\leq \sup_{x, y \in (BV), (x, y) > 0} M \int_0^1 |x - y| dt \\ &\quad + \left[ \int_0^1 |x - y| dt + |L_0^1(x) - L_0^1(y)| \right] \leq M. \end{aligned}$$

Since  $(CBV)$  and  $(AC)$  are dense in  $(BV)$ , a functional  $f$  additive and uniformly continuous on either subspace can be extended to be uniformly continuous on  $(BV)$ . As has been done in the proof of Theorem 5.3 of A, one may then show that the extended functional is additive on  $(BV)$ . Consequently the form of  $f$  is determined as asserted.

That the general form of the continuous additive functional on  $(BV)$  is (3) can be shown by essentially the same argument as has been used in the proof of Theorem 3 above.

To determine the form of the continuous additive functional on  $(CBV)$ , it should be noted that if the set  $P$  in Lemma 2 is of measure zero,  $x$  is a singular function. According to a theorem of Morse<sup>(26)</sup>, since  $x$  is singular,  $(x_n, x) \rightarrow 0$  in the metric (1) implies  $(x_n, x) \rightarrow 0$  in the metric (17); hence functions  $\lambda, \mu$  exist such that in the metric (17),  $(\lambda, \mu)$  is  $< \delta$  and  $\int_0^1 h(t)\lambda'(t)dt - \int_0^1 h(t)\mu'(t)dt$  is  $\geq k/2$ . Since any non-vacuous perfect set contains such a set with measure zero, only trivial modifications in the proof of Theorem 4 need now be made in order to establish the desired result.

Since in  $(AC)_0$  convergence in the metric (17) is equivalent<sup>(27)</sup> to convergence in the metric  $(x, y) = T_0^1(x - y)$ , it follows that the general form of the continuous additive functional on  $(AC)_0$  is (6) and on  $(AC)$  is (8).

The converse statements concerning the uniform continuity of (2) on  $(BV)$  and the continuity of (3), (4), and (8) on  $(BV)$ ,  $(CBV)$ , and  $(AC)$  respectively are very easily verified.

The fact that, if  $x_1$  is an arbitrary point of  $(AC)$ , the functional (8) on  $(AC)$  satisfies a Lipschitz condition at  $x_1$  only when  $h(t)$ , and therefore  $g(t) = h(0) - h(t)$ , satisfies a Lipschitz condition on  $I$  is a consequence of

**LEMMA 3.** *The functional (8) has the following property when  $x_1$  is any point of  $(AC)$ :*

<sup>(26)</sup> See Morse, *Convergence in variation and related topics*, these Transactions, vol. 41 (1937), pp. 48-83, Theorem 5.2.

<sup>(27)</sup> The equivalence is a consequence of Theorems 4 and 5 of Adams and Lewy, loc. cit.



$$(18) \quad \sup_{x \in (AC), (x, x_1) > 0} |f(x) - f(x_1)| / (x, x_1) \\ \geq \sup_{0 \leq t_1 < t_2 \leq 1} |h(t_1) - h(t_2)| / |t_1 - t_2|.$$

**Proof.** In the light of remarks made in the paragraph following (7) it suffices to prove that the left member of (18) is  $\geq |h(b) - h(a)| / (b - a)$ , where  $0 \leq a < b \leq 1$ , in four cases: (i) that in which  $a = 0$ ,  $b = 1$ , which is trivially verified by taking  $x(t) = x_1(t) + 1$  for  $t \in I$ ; (ii) that in which  $0 < a < b < 1$  with  $a$  and  $b$  points of the Lebesgue set<sup>(28)</sup> for the function  $h$ ; (iii) that in which  $a = 0$  and  $b$ ,  $0 < b < 1$ , is a point of the Lebesgue set for  $h$ ; and (iv) that in which  $b = 1$  and  $a$ ,  $0 < a < 1$ , is a point of the Lebesgue set for  $h$ .

To dispose of case (ii) let  $0 < \delta < (b - a)/2$ , let  $\epsilon$  be an arbitrary number (positive, negative, or zero), and consider the function

$$x_{\delta, \epsilon}(t) = \begin{cases} x_1(t) & \text{for } 0 \leq t \leq a, b \leq t \leq 1, \\ x_1(t) + \epsilon(t - a)/\delta & \text{for } a \leq t \leq a + \delta, \\ x_1(t) + \epsilon & \text{for } a + \delta \leq t \leq b - \delta, \\ x_1(t) - \epsilon(t - b)/\delta & \text{for } b - \delta \leq t \leq b. \end{cases}$$

For each  $\delta$  and  $\epsilon$  we have

$$\begin{aligned} \int_0^1 |x_{\delta, \epsilon}(t) - x_1(t)| dt &= |\epsilon| \cdot (b - a - \delta), \\ |f(x_{\delta, \epsilon}) - f(x_1)| &= |f(x_{\delta, \epsilon}) - f(x_1)| = \left| \int_0^1 h(t) [x'_{\delta, \epsilon}(t) - x'_1(t)] dt \right| \\ &= \left| \int_a^{a+\delta} h(t) \epsilon/\delta dt - \int_{b-\delta}^b h(t) \epsilon/\delta dt \right| \\ &= \left| \epsilon [h(a) - h(b)] + \epsilon \int_a^{a+\delta} [h(t) - h(a)] dt/\delta \right. \\ &\quad \left. - \epsilon \int_{b-\delta}^b [h(t) - h(b)] dt/\delta \right| \\ &\geq |\epsilon| \cdot |h(b) - h(a)| - |\epsilon| \eta_1(\delta) - |\epsilon| \eta_2(\delta), \end{aligned}$$

where, as  $\delta \rightarrow 0$ ,

$$(19) \quad \begin{aligned} \eta_1(\delta) &= \int_a^{a+\delta} |h(t) - h(a)| dt/\delta \rightarrow 0, \\ \eta_2(\delta) &= \int_{b-\delta}^b |h(t) - h(b)| dt/\delta \rightarrow 0. \end{aligned}$$

<sup>(28)</sup> See, for example, Titchmarsh, *The Theory of Functions*, Oxford, 1932, p. 364. For each  $h$  summable on  $I$  the "Lebesgue set" of points  $t$  where  $\int_t^{t+\delta} |h(u) - h(t)| du/\delta$  tends to zero with  $\delta$  has measure 1.

For each fixed  $\delta$ , the function

$$\begin{aligned}\phi(\delta, \epsilon) &= L_0^1(x_{\delta, \epsilon}) - L_0^1(x_1) \\ &= \int_a^{a+\delta} \{1 + [x_1'(t) + \epsilon]^2\}^{1/2} dt + \int_{b-\delta}^b \{1 + [x_1'(t) - \epsilon]^2\}^{1/2} dt \\ &\quad - L_a^{a+\delta}(x_1) - L_{b-\delta}^b(x_1)\end{aligned}$$

vanishes at  $\epsilon=0$ , and  $\partial\phi/\partial\epsilon$  may easily be seen to exist<sup>(29)</sup> for each  $\epsilon$ . If  $\partial\phi/\partial\epsilon=0$  at  $\epsilon=0$ , we have

$$(20) \quad \frac{|f(x_{\delta, \epsilon}) - f(x_1)|}{(x_{\delta, \epsilon}, x_1)} \geq \frac{|h(b) - h(a)| - \eta_1(\delta) - \eta_2(\delta)}{b - a - \delta + |\phi(\delta, \epsilon) - \phi(\delta, 0)|/\epsilon},$$

with  $[\phi(\delta, \epsilon) - \phi(\delta, 0)]/\epsilon \rightarrow 0$  as  $\epsilon \rightarrow 0$ . If  $\partial\phi/\partial\epsilon \neq 0$  at  $\epsilon=0$ , there exists a unilateral neighborhood of zero such that for  $\epsilon$  in this neighborhood,  $L_0^1(x_{\delta, \epsilon})$  is  $< L_0^1(x_1)$ . Moreover it is evident that  $L_0^1(x_{\delta, \epsilon})$  is a continuous function of  $\epsilon$  which becomes positively infinite as  $\epsilon \rightarrow +\infty$  or  $\epsilon \rightarrow -\infty$ . Hence there exists an  $\epsilon$ , say  $\epsilon_\delta$ , for which  $L_0^1(x_{\delta, \epsilon_\delta}) = L_0^1(x_1)$ . In fact the inequality<sup>(30)</sup>

$$\begin{aligned}L_a^{a+\delta}(x_{\delta, \epsilon}) + L_{b-\delta}^b(x_{\delta, \epsilon}) &> T_a^{a+\delta}(x_{\delta, \epsilon}) + T_{b-\delta}^b(x_{\delta, \epsilon}) \\ &\geq T_a^{a+\delta}(x_{\delta, \epsilon} - x_1) - T_a^{a+\delta}(x_1) \\ &\quad + T_{b-\delta}^b(x_{\delta, \epsilon} - x_1) - T_{b-\delta}^b(x_1) \\ &= 2|\epsilon| - T_a^{a+\delta}(x_1) - T_{b-\delta}^b(x_1)\end{aligned}$$

shows that for  $|\epsilon| = L_a^{a+\delta}(x_1) + L_{b-\delta}^b(x_1) - L_0^1(x_1)$ ,  $L_0^1(x_{\delta, \epsilon})$  exceeds  $L_0^1(x_1)$ ; this gives an upper bound for  $|\epsilon_\delta|$  and shows incidentally that  $\epsilon_\delta \rightarrow 0$  with  $\delta$ . Choosing  $\epsilon = \epsilon_\delta$ , we have

$$(21) \quad \frac{|f(x_{\delta, \epsilon}) - f(x_1)|}{(x_{\delta, \epsilon}, x_1)} \geq \frac{|h(b) - h(a)| - \eta_1(\delta) - \eta_2(\delta)}{b - a - \delta}.$$

The inequality (18) now follows at once from (19), (20), and (21).

Although cases (iii) and (iv) are not formally symmetric, it should suffice to examine one of them. In case (iii), for example, we may define

$$x_{\delta, \epsilon}(t) = \begin{cases} x_1(t) & \text{for } b \leq t \leq 1, \\ x_1(t) - \epsilon(t - b)/\delta & \text{for } b - \delta \leq t \leq b, \\ x_1(t) + \epsilon & \text{for } 0 \leq t \leq b - \delta, \end{cases}$$

and find

<sup>(29)</sup> See, for example, Hobson, loc. cit., 2d edition, vol. 2, Cambridge, 1926, p. 355.

<sup>(30)</sup> See Adams and Lewy, loc. cit., inequalities (2) and (3).

$$\begin{aligned}
\int_0^1 |x_{\delta,\epsilon}(t) - x_1(t)| dt &= |\epsilon| \cdot (b - \delta/2), \\
|f(x_{\delta,\epsilon}) - f(x_1)| &= \left| \epsilon h(0) + \int_0^1 h(t) [x'_{\delta,\epsilon}(t) - x'_1(t)] dt \right| \\
&= \left| \epsilon h(0) - \int_{b-\delta}^b h(t) \epsilon / \delta dt \right| \\
&= \left| \epsilon [h(0) - h(b)] - \epsilon \int_{b-\delta}^b [h(t) - h(b)] dt / \delta \right| \\
&\geq |\epsilon| \cdot |h(b) - h(0)| - |\epsilon| \eta_2(\delta),
\end{aligned}$$

whence we may proceed as before in case (ii).

To complete the proof of the results stated in the theorem, it is sufficient to make the following remark. Let  $f$  be a continuous additive functional on  $(BV) [(CBV)]$ , and let  $y_1$  be an arbitrary point of  $(BV) [(CBV)]$ . Set

$$x_1(t) = \int_0^t y'_1(u) du, \quad z_1(t) = y_1(t) - x_1(t), \quad t \in I,$$

so that  $x_1$  is the absolutely continuous part, and  $z_1$  the singular part, of the function  $y_1$ . Then we have, for  $y \in (BV) [y \in (CBV)]$ ,

$$\begin{aligned}
\sup_{(y, y_1) > 0} |f(y - y_1)| / \Lambda(y, y_1) \\
&\geq \sup_{x \in (AC), (x, z_1) > 0} |f(x + z_1 - x_1 - z_1)| / (x + z_1, x_1 + z_1) \\
&= \sup_{x \in (AC), (x, z_1) > 0} |f(x - x_1)| / (x, x_1),
\end{aligned}$$

$$\text{since } |L_0^1(x + z_1) - L_0^1(x_1 + z_1)| = |L_0^1(x) + T_0^1(z_1) - L_0^1(x_1) - T_0^1(z_1)|.$$

BROWN UNIVERSITY,  
PROVIDENCE, R. I.,  
THE UNIVERSITY OF CALIFORNIA,  
BERKELEY, CALIF.

## A NEW SPECIAL FORM OF THE LINEAR ELEMENT OF A SURFACE

BY  
JESSE DOUGLAS

**1. Introduction and statement of results.** The great circles of a sphere form a family of  $\infty^2$  curves having the following three important properties:

- (1) They are *geodesics* of the sphere.
- (2) They are a *linear* system; that is, a point transformation exists which converts them into the straight lines of a plane. Indeed, central projection of the sphere on any plane not passing through its center will accomplish such a transformation. An equivalent statement is that it is possible to introduce coordinates  $u, v$  on the spherical surface so that the totality of great circles is represented by the general linear equation:  $au + bv + c = 0$ .
- (3) *The angular excess of any triangle ABC formed by great circles is proportional to the area of the triangle:*

$$(1.1) \quad \mathcal{E} = A + B + C - \pi = kA,$$

where the factor of proportionality  $k$  is equal to the Gaussian curvature of the sphere:  $k = 1/R^2$ .

It is evident that all the geometric entities and properties involved in these three statements are invariant under any *bending* or *isometric transformation* of the spherical surface<sup>(1)</sup> together with its great circles; this means a point transformation into a family of  $\infty^2$  curves upon another surface so that  $ds = ds'$ , where  $ds$  denotes the element of length of the sphere and  $ds'$  the corresponding element of the transformed surface. According to a fundamental theorem of Gauss, the Gaussian curvature  $K$  of the transformed surface must be the same as that of the sphere, therefore constant. Evidently, too, the geodesics of the sphere go over into the geodesics of the transformed surface. It follows that *the three stated properties are possessed also by the geodesics of any surface of constant Gaussian curvature.*

Let us denote by  $(\mathcal{Y}, S)$  the geometric configuration consisting of a family  $\mathcal{Y}$  of  $\infty^2$  curves upon a surface  $S$ . Then it is obvious that, in respect of the possession or non-possession of any of our three properties, the configuration  $(\mathcal{Y}, S)$  is completely equivalent to any configuration  $(\mathcal{Y}', S')$  derived therefrom by isometric transformation. Any two such isometric configurations will

---

Presented to the Society, April 27, 1940; received by the editors March 5, 1940. This paper was received by the editors of the *Annals of Mathematics*, May 11, 1939, accepted by them, and later transferred to these *Transactions*.

(<sup>1</sup>) Meaning a properly limited region of the spherical surface. As is well known, the sphere as a whole is indeformable. In general, all our considerations will be local or differential-geometric.

therefore be regarded as essentially identical. In other words, all that is relevant concerning the surface  $S$  is its first fundamental form

$$(1.2) \quad ds^2 = Edu^2 + 2Fdudv + Gdv^2.$$

The family  $\mathcal{F}$  may always be defined by a differential equation of second order:

$$(1.3) \quad v'' = \phi(u, v, v') \quad (v' = dv/du, v'' = d^2v/du^2).$$

Thus any configuration  $(\mathcal{F}, S)$  is represented analytically by a system of functions  $[E(u, v), F(u, v), G(u, v), \phi(u, v, v')]$ .

With every two properties that may be selected from the three stated at the beginning, we may associate a corresponding converse problem. Thus we may ask for all configurations  $(\mathcal{F}, S)$  which have:

- (a) properties (1) and (3),
- (b) properties (1) and (2),
- (c) properties (2) and (3).

The answer to the converse question (a) is classical. According to a theorem of Gauss, the angular excess  $\mathcal{E}$  of any triangle  $ABC$  formed by three geodesics of a surface  $S$  is given by the formula<sup>(2)</sup>

$$\mathcal{E} = \iint K d\omega \quad \text{over } ABC,$$

where  $d\omega$  denotes the element of area. By the law of the mean, this gives  $\mathcal{E} = K(m)\mathcal{A}$ , where  $K(m)$  denotes the value of the Gaussian curvature at some point  $m$  of  $ABC$ , while  $\mathcal{A}$  denotes the area of this triangle. It follows immediately that, as the triangle  $ABC$  shrinks to any fixed point  $p$  of  $S$ ,

$$\lim \mathcal{E}/\mathcal{A} = K(p).$$

Property (3) then implies that the Gaussian curvature of the surface  $S$  is constant:  $K(p) = k$  for every point  $p$  of  $S$ . That the family  $\mathcal{F}$  consists of the geodesics of  $S$  is part of the data of problem (a).

The answer to the converse question (b) is also classical, having been given by Beltrami in 1865<sup>(3)</sup>. He proved the theorem: *if a surface  $S$  can be represented point by point on a plane so that the geodesics of  $S$  correspond to the straight lines of the plane, then  $S$  has constant Gaussian curvature*. Thus, again, the only solution of the converse problem is the one which is known *a priori*.

*The converse problem (c).* It is curious that the converse problem (c) has not hitherto been studied. Here I have found the solution  $(\mathcal{F}, S)$  to be *more general* than the geodesics of a surface of constant curvature. In fact, the

<sup>(2)</sup> In formula (2.2) of the next section, let  $1/\rho = 0$ , expressing the geodesic character of the sides of the triangle.

<sup>(3)</sup> E. Beltrami, *Opere Matematiche*, vol. 1, pp. 262-280.

complete solution is given by the formulas which follow, whose derivation constitutes the purpose of the present paper.

In formulating our results, it is convenient to use—instead of general coordinates  $u, v$  upon the surface  $S$ , wherein  $ds^2$  has the form (1.2)—minimal coordinates, wherein

$$(1.4) \quad ds^2 = 2F du dv.$$

The characteristic property of minimal coordinates is that the coordinate lines  $u = \text{const.}$ ,  $v = \text{const.}$  have zero length<sup>(4)</sup>, or are the minimal lines of the surface  $S$ . Such coordinates are determined only up to an arbitrary transformation  $u_1 = \phi(u)$ ,  $v_1 = \psi(v)$ , which preserves the minimal character of the coordinate lines.

Let  $U_1, U_2$  denote any two functions of  $u$  alone, and  $V_1, V_2$  any two functions of  $v$  alone. Form the determinants

$$(1.5) \quad I = \begin{vmatrix} U_1 + V_1 & U_1' \\ U_2 + V_2 & U_2' \end{vmatrix}, \quad II = \begin{vmatrix} U_1 + V_1 & V_1' \\ U_2 + V_2 & V_2' \end{vmatrix},$$

where the accent denotes differentiation. We always suppose  $U_1, U_2, V_1, V_2$  so chosen that  $I \neq 0$ ,  $II \neq 0$ . Then we shall prove that the most general configuration  $(\mathcal{F}, S)$  having the properties (2), (3) is represented by the formulas:

$$(1.6) \quad \mathcal{F}: v'' = Bv' + Cv'^2$$

where

$$(1.6') \quad B = \frac{\partial}{\partial u} (\log I - 2 \log II), \quad C = \frac{\partial}{\partial v} (2 \log I - \log II),$$

while, for the surface  $S$ ,

$$(1.7) \quad S: 2kF = \frac{\partial^2}{\partial u \partial v} (\log I + \log II).$$

With the help of a simple determinant transformation<sup>(5)</sup>, we find

$$(1.8) \quad \frac{\partial^2}{\partial u \partial v} \log I = \frac{II}{I^2} \begin{vmatrix} U_1' & U_1'' \\ U_2' & U_2'' \end{vmatrix}, \quad \frac{\partial^2}{\partial u \partial v} \log II = \frac{I}{II^2} \begin{vmatrix} V_1' & V_1'' \\ V_2' & V_2'' \end{vmatrix};$$

hence the expanded form of (1.7) is

$$(1.7') \quad 2kF = \frac{II}{I^2} \begin{vmatrix} U_1' & U_1'' \\ U_2' & U_2'' \end{vmatrix} + \frac{I}{II^2} \begin{vmatrix} V_1' & V_1'' \\ V_2' & V_2'' \end{vmatrix}.$$

*The case  $k=0$ .* In interpreting these results, it is important to distinguish

<sup>(4)</sup> Of course, these coordinate lines must be imaginary.

<sup>(5)</sup> Formula (3.19).



the case  $k=0$  from  $k \neq 0$ . If  $k=0$ , property (3) becomes the statement that the sum of the angles of every triangle of  $\mathcal{F}$  is two right angles. In addition,  $\mathcal{F}$  is required to be linear, by property (2). Now, an arbitrary surface  $S$  is capable of conformal representation upon a plane  $P$ , and in infinitely many ways, since we may combine any given conformal representation of  $S$  on  $P$  with an arbitrary conformal transformation of  $P$  into itself:  $u' + iv' = f(u + iv)$ . Let  $\mathcal{F}$  denote the family of  $\infty^2$  curves on  $S$  which results from the family of all straight lines of the plane  $P$  by any conformal map of  $S$  on  $P$ . Then  $\mathcal{F}$  is obviously linear, and also the angle-sum of every triangle of  $\mathcal{F}$  is two right angles, since these are conformally invariant properties which belong to the straight lines. Thus, an arbitrary surface  $S$  carries infinitely many families of curves  $\mathcal{F}$  which are linear and in which every triangle has an angle-sum of two right angles.

This finds expression in the formula (1.7') by the circumstance that when  $k=0$  this formula implies no restriction on the function  $F$ , that is, on the  $ds^2$  of  $S$ , but rather only a condition on the functions  $U_1, U_2, V_1, V_2$  which determine the family  $\mathcal{F}$ . Indeed, if  $k=0$ , it follows from (1.7) that

$$(1.9) \quad I \cdot II = U_3 V_3,$$

and from (1.7') that either

$$(1.9') \quad \frac{I}{II} = \frac{U_4}{V_4}$$

where

$$U_4 = \left| \begin{array}{cc} U_1' & U_1'' \\ U_2' & U_2'' \end{array} \right|^{1/2}, \quad V_4 = - \left| \begin{array}{cc} V_1' & V_1'' \\ V_2' & V_2'' \end{array} \right|^{1/2},$$

or else that

$$(1.10) \quad \left| \begin{array}{cc} U_1' & U_1'' \\ U_2' & U_2'' \end{array} \right| = 0, \quad \left| \begin{array}{cc} V_1' & V_1'' \\ V_2' & V_2'' \end{array} \right| = 0.$$

By (1.9) and (1.9'),

$$(1.11) \quad I = U_3 V_3, \quad II = U_3 V_3;$$

therefore

$$\frac{\partial^2}{\partial u \partial v} \log I = 0, \quad \frac{\partial^2}{\partial u \partial v} \log II = 0,$$

whence by (1.8), since by hypothesis  $I \neq 0, II \neq 0$ , we deduce

$$(1.12) \quad \left| \begin{array}{cc} U_1' & U_1'' \\ U_2' & U_2'' \end{array} \right| = 0, \quad \left| \begin{array}{cc} V_1' & V_1'' \\ V_2' & V_2'' \end{array} \right| = 0.$$

(1.10) is the same as (1.12), which therefore subsists in either case. (1.12) implies the existence of linear relations with constant coefficients:

$$(1.13) \quad c_1 U_1 + c_2 U_2 = c_3, \quad c'_1 V_1 + c'_2 V_2 = c'_3,$$

where either  $c_1$  or  $c_2 \neq 0$  and either  $c'_1$  or  $c'_2 \neq 0$ .

It is evident by the defining formulas (1.5) that, under the conditions (1.13), the functions I, II must have the forms (1.11). Therefore, by (1.6'),

$$B = U, \quad C = V;$$

accordingly, the differential equations (1.6) of  $\mathcal{F}$  have the form

$$(1.14) \quad v'' = Uv' + Vv'^2.$$

It is easily verified that this is the general form of differential equation for a family  $\mathcal{F}$  derivable by conformal transformation:  $u_1 = \phi(u)$ ,  $v_1 = \psi(v)$ , from the straight lines of a plane:  $v_1'' = 0$ —explicitly,  $U = \phi''(u)/\phi'(u)$ ,  $V = -\psi''(v)/\psi'(v)$ . In summary, we have a proof of the following theorem:

*If a family  $\mathcal{F}$  of  $\infty^2$  curves on a surface  $S$  is linear, and the sum of the angles in every triangle of  $\mathcal{F}$  is two right angles, then  $\mathcal{F}$  must be a conformal image of the  $\infty^2$  straight lines of a plane.*

In a previous paper<sup>(6)</sup>, the author proved this theorem synthetically. The first statement and proof is an analytic one by E. Kasner<sup>(7)</sup>.

The case  $k \neq 0$ . Of more interest is the case  $k \neq 0$ . Then formula (1.7) or (1.7') really specializes the surface  $S$ : its  $ds^2$  must have, in minimal coordinates, the form

$$(1.15) \quad ds^2 = \frac{1}{k} \frac{\partial^2}{\partial u \partial v} (\log I + \log II) du dv \\ = \frac{1}{k} \left\{ \frac{II}{I^2} \begin{vmatrix} U_1' & U_1'' \\ U_2' & U_2'' \end{vmatrix} + \frac{I}{II^2} \begin{vmatrix} V_1' & V_1'' \\ V_2' & V_2'' \end{vmatrix} \right\} du dv.$$

Upon all and only such surfaces  $S$  can curve families  $\mathcal{F}$  be found with properties (2) and (3).

In order that this form of the surface  $S$  shall not be degenerate, it is necessary and sufficient (besides  $I \neq 0$ ,  $II \neq 0$ ) that  $F \neq 0$ , or

$$\frac{\partial^2}{\partial u \partial v} (\log I + \log II) \neq 0.$$

But in the contrary case, we have seen by the calculations (1.9)–(1.13) that we must have (1.12) or its equivalent (1.13). Conversely, it is evident that (1.12) implies  $F = 0$ .

<sup>(6)</sup> Number 2 of the list of references at the end.

<sup>(7)</sup> Reference [1].

Hence, under the hypothesis  $I \neq 0$ ,  $II \neq 0$ , the formula (1.15) will give a nondegenerate surface  $S$  when and only when linear relations of the form (1.13) do not subsist simultaneously between  $U_1, U_2$  and between  $V_1, V_2$ . If  $U_1, U_2$  are not both constant and  $V_1, V_2$  are not both constant, the condition that relations of the form (1.13) or (1.12) shall not hold simultaneously is sufficient to guarantee in addition that  $I \neq 0$ ,  $II \neq 0$ .

This completes our description of the special form of the linear element of the surface  $S$  signified by the title of the present paper.

An indication that this type of surface  $S$  is more general than one of constant curvature results by a count of arbitrary functions. The most general form of the  $ds^2$  of a surface of constant curvature  $c$  referred to minimal coordinates is<sup>(8)</sup>

$$(1.16) \quad ds^2 = \frac{4U'V'}{c(U-V)^2} du dv,$$

thus involving, besides the arbitrary constant  $c$ , only the *two* arbitrary functions  $U$  of  $u$  and  $V$  of  $v$ , which determine the distribution of parametric values  $u, v$  over the two systems of minimal lines respectively. The formula (1.15), on the other hand, involves *four* general functions  $U_1, U_2, V_1, V_2$ , subject only to the slight restrictions of linear independence which we have mentioned. Of these four functions, two correspond to an arbitrary transformation  $u_1 = \phi(u)$ ,  $v_1 = \psi(v)$  on the surface  $S$  which conserves the minimal lines, so that only two of the arbitrary functions are really effective in varying the form of  $S$ . We may say that, if isometric surfaces are regarded as identical, there are only  $\infty^1$  surfaces of constant curvature, depending on the value  $c$  of this curvature, whereas the category of surfaces  $S$  with properties (2) and (3) involves two arbitrary functions of one variable.

This indication is, of course, not completely decisive, since there remains the question of whether  $U_1, U_2, V_1, V_2$  are all essential. To obtain a definite proof that the formula (1.7') contains surfaces *not* of constant curvature, we may calculate the Gaussian curvature by the formula

$$(1.17) \quad K = -\frac{1}{F} \frac{\partial^2}{\partial u \partial v} \log F = \frac{F_u F_v - F F_{uv}}{F^3}.$$

The result is a rational function of  $U_1, U_2, V_1, V_2$  and their derivatives of the first three orders. A partial calculation suffices to show that this rational function does not reduce identically to a constant when all the quantities mentioned are considered as independent variables—as they may be, since the functions  $U_1, U_2, V_1, V_2$  are arbitrary, and they and any finite number of their derivatives are therefore capable of taking any assigned values for any finite number of given values of  $(u, v)$ . Consequently, we can arrange to give

<sup>(8)</sup> Cf. G. Darboux, *Théorie Générale des Surfaces*, 1887 edition, vol. 1, p. 30. Write  $x = U$ ,  $y = V$  in formula (1), p. 30.

these functions and those of their derivatives which appear in the expression for  $K$  such particular values at any two chosen points  $(u_1, v_1)$ ,  $(u_2, v_2)$  that  $K(u_1, v_1) \neq K(u_2, v_2)$ ; therefore  $K$  will not be a constant.

2. **Conditions for the property  $\mathcal{E} = k\mathcal{A}$ .** We begin the proof of the results stated in §1 by recalling the formula of Gauss-Bonnet<sup>(9)</sup>. If  $\Gamma$  denote any closed curve with continuously turning tangent, bounding a region  $R$ , then

$$\int_{\Gamma} \frac{ds}{\rho} + \iint_R K d\omega = 2\pi,$$

where  $1/\rho$  is the geodesic curvature of  $\Gamma$ . In case the curve  $\Gamma$  has corners at the points  $P_i$  ( $i=1, 2, \dots, m$ ), then this formula must be modified as follows:

$$(2.1) \quad \int_{\Gamma} \frac{ds}{\rho} + \iint_R K d\omega + \sum_{i=1}^m \theta_i = 2\pi.$$

Here  $\theta_i$  represents the angle, taken with proper sign, between the sensed tangents to the two arcs of  $\Gamma$  which form the corner at  $P_i$ .

Let us apply formula (2.1) to any triangle  $ABC$  formed by three curves of  $\mathcal{F}$ . The boundary  $\Gamma$  of this triangle has corners at  $P_1, P_2, P_3 = A, B, C$ , and  $\theta_1 = \pi - A$ ,  $\theta_2 = \pi - B$ ,  $\theta_3 = \pi - C$ , where  $A, B, C$  denote the interior angles of the triangle. Consequently, by substitution in (2.1),

$$(2.2) \quad \int_{\Gamma} \frac{ds}{\rho} + \iint_{ABC} K d\omega = A + B + C - \pi = \mathcal{E}.$$

By property (3),

$$(2.3) \quad \mathcal{E} = k\mathcal{A} = \iint_{ABC} k d\omega;$$

therefore

$$(2.4) \quad \int_{\Gamma} \frac{ds}{\rho} = \iint_{ABC} (k - K) d\omega = \iint_{ABC} (k - K) W du dv,$$

since

$$(2.5) \quad d\omega = W du dv,$$

where  $W = (EG - F^2)^{1/2}$ .

Every polygon whose sides are curves of  $\mathcal{F}$  can be decomposed into triangles of  $\mathcal{F}$ . It follows, by the additive nature of both contour and surface integration, that the formula (2.4) applies to any polygon of  $\mathcal{F}$ ; that is, if  $\Gamma$  denotes the boundary and  $R$  the interior of any polygon of  $\mathcal{F}$ , we have

$$(2.6) \quad \int_{\Gamma} \frac{ds}{\rho} = \iint_R (k - K) W du dv.$$

<sup>(9)</sup> See W. Blaschke, *Vorlesungen über Differentialgeometrie*, 1921, p. 108.

By Green's theorem, the surface integral over  $R$  can be expressed as a contour integral over  $\Gamma$ :

$$(2.7) \quad \iint_R (k - K)W du dv = \int_{\Gamma} P_1 du + Q_1 dv,$$

where  $P_1, Q_1$  are any two functions of  $u, v$  which obey

$$(2.8) \quad \frac{\partial Q_1}{\partial u} - \frac{\partial P_1}{\partial v} = (k - K)W$$

—the existence of such functions is obvious.

Relations (2.6) and (2.7) give

$$(2.9) \quad \int_{\Gamma} \left( \frac{ds}{\rho} - P_1 du - Q_1 dv \right) = 0$$

for every polygon  $\Gamma$  of  $F$ . This implies that the same integral taken over any polygonal path between any two fixed points of  $S$  does not depend on the path itself but only on its end-points. (By a "polygonal path" we mean one composed of a finite number of arcs of curves of  $F$ .) According to a standard theorem<sup>(10)</sup>, it follows that the element of integration in (2.9) is an exact differential:

$$\frac{ds}{\rho} - P_1 du - Q_1 dv = \lambda_u du + \lambda_v dv,$$

where the subscripts denote partial differentiation of the arbitrary function  $\lambda(u, v)$ . Thus we have

$$(2.10) \quad \frac{ds}{\rho} = P du + Q dv,$$

where  $P = P_1 + \lambda_u$ ,  $Q = Q_1 + \lambda_v$ . Obviously,  $P, Q$  also obey the condition (2.8):

$$(2.11) \quad Q_u - P_v = (k - K)W,$$

since  $\lambda_{uv}$  cancels in the process of substitution.

Conversely, it is easily seen that if (2.10) is obeyed along every curve of a family  $\mathcal{F}$ , and  $P, Q$  are related by (2.11), then  $\mathcal{F}$  has property (3), as expressed by (2.3).

Formula (2.10) by itself defines a type of curve family called a *velocity family*<sup>(11)</sup>. Thus property (3) is characteristic of a particular kind of velocity family—one where formula (2.11) is obeyed.

<sup>(10)</sup> It is easily seen to be sufficient for the application of this theorem that the condition of independence of the path of integration apply merely to polygonal paths of  $\mathcal{F}$ .

<sup>(11)</sup> The name is due to E. Kasner, these Transactions, vol. 10 (1909), p. 213. The geodesics of a Weyl metric are a general velocity family; see H. Weyl, *Raum, Zeit, Materie*, 3d edition,

In minimal coordinates, where  $ds^2 = 2Fdu dv$ , the geodesic curvature  $1/\rho$  of any curve  $v=v(u)$  is given by<sup>(12)</sup>

$$(2.12) \quad \frac{ds}{\rho} = \frac{-v'' + (F_u/F)v' - (F_v/F)v'^2}{2iv'} du.$$

Therefore, for a velocity family, we have by (2.10):

$$(2.13) \quad v'' = Bv' + Cv'^2,$$

where

$$(2.14) \quad \begin{aligned} B &= F_u/F - 2iP = (\log F)_u - 2iP, \\ C &= -F_v/F - 2iQ = -(\log F)_v - 2iQ. \end{aligned}$$

Conversely, a family  $\mathcal{F}$  whose differential equation in minimal coordinates is of the type (2.13), where  $B, C$  are any functions of  $u, v$ , obeys (2.10) with  $P, Q$  defined by (2.14). That is: *the form (2.13) is characteristic of velocity families.*

Let us now apply the condition (2.11) by calculating

$$(2.15) \quad C_u - B_v = -2(\log F)_{uv} - 2i(Q_u - P_v).$$

In minimal coordinates, where  $E=0, G=0, W=(EG-F^2)^{1/2}=iF$ , the Gaussian curvature  $K$  is expressed by (1.17), so that the condition (2.11) becomes

$$(2.16) \quad Q_u - P_v = ikF + i(\log F)_{uv}.$$

This gives, when substituted in (2.15),

$$(2.17) \quad C_u - B_v = 2kF.$$

Conversely, (2.17) gives (2.16) when substituted in the identity (2.15).

In summary: *property (3) is expressed in minimal coordinates by the formulas (2.13), (2.17).*

**3. Linearity of the family  $\mathcal{F}$ .** We now have to impose the additional property (2) of linearity on the family  $\mathcal{F}$ .

If we apply an arbitrary coordinate transformation

$$(3.1) \quad u_1 = \phi(u, v), \quad v_1 = \psi(u, v)$$

to  $\mathcal{F}$ , the effect on the derivatives  $v', v''$  is as follows:

$$(3.2) \quad v'_1 = \frac{\psi_u + \psi_v v'}{\phi_u + \phi_v v'},$$

1919, p. 112. Cf. also C. H. Rowe [4]. The term "zyklisches Netz" used by J. Radon, following Blaschke, denotes the same thing as a velocity family; see J. Radon [3].

<sup>(12)</sup> Blaschke, loc. cit., p. 117. Write  $E=0, G=0, u'=1, u''=0$ .



$$(3.3) \quad v_1'' = \{(\phi_u + \phi_v v')(\psi_{uu} + 2\psi_{uv}v' + \psi_{vv}v'^2 + \psi_v v'') - (\psi_u + \psi_v v')(\phi_{uu} + 2\phi_{uv}v' + \phi_{vv}v'^2 + \phi_v v'')\} / (\phi_u + \phi_v v')^3.$$

Suppose that after this transformation the finite equation of  $\mathcal{F}$  has the linear form  $v_1 = au_1 + b$ , or the differential equation of  $\mathcal{F}$  becomes  $v_1'' = 0$ . Then in the original coordinate system  $(u, v)$  the differential equation of  $\mathcal{F}$  is, by (3.3),

$$(3.4) \quad v'' = A + Bv' + Cv'^2 + Dv'^3$$

where

$$(3.5) \quad \begin{aligned} A &= \frac{\psi_u \phi_{uu} - \phi_u \psi_{uu}}{\Delta}, \\ B &= 2 \frac{\psi_u \phi_{uv} - \phi_u \psi_{uv}}{\Delta} + \frac{\psi_v \phi_{uu} - \phi_v \psi_{uu}}{\Delta}, \\ C &= 2 \frac{\psi_v \phi_{uv} - \phi_v \psi_{uv}}{\Delta} + \frac{\psi_u \phi_{vv} - \phi_u \psi_{vv}}{\Delta}, \\ D &= \frac{\psi_v \phi_{vv} - \phi_v \psi_{vv}}{\Delta}, \\ \Delta &= \phi_u \psi_v - \psi_u \phi_v \neq 0. \end{aligned}$$

This is, consequently, the general form of the differential equation of a linear family in any system of coordinates.

If now the coordinates are minimal, then the necessary and sufficient condition for  $\mathcal{F}$  to be a velocity family is, by comparison of (3.4) with (2.13),  $A=0, D=0$ , that is,

$$(3.6) \quad \psi_u \phi_{uu} - \phi_u \psi_{uu} = 0, \quad \psi_v \phi_{vv} - \phi_v \psi_{vv} = 0.$$

This gives

$$(3.7) \quad \frac{\phi_{uu}}{\phi_u} = \frac{\psi_{uu}}{\psi_u} = \rho, \quad \frac{\phi_{vv}}{\phi_v} = \frac{\psi_{vv}}{\psi_v} = \sigma,$$

or

$$(3.8) \quad \rho = (\log \phi_u)_u = (\log \psi_u)_u, \quad \sigma = (\log \phi_v)_v = (\log \psi_v)_v.$$

We calculate

$$(\log \Delta)_u = \frac{\Delta_u}{\Delta} + \frac{(\phi_u \psi_{uv} - \psi_u \phi_{uv}) + (\phi_{uv} \psi_v - \psi_{uv} \phi_v)}{\Delta} = \frac{\phi_u \psi_{uv} - \psi_u \phi_{uv}}{\Delta} + \rho,$$

similarly

$$(\log \Delta)_v = \frac{\Delta_v}{\Delta} = \frac{\psi_v \phi_{uv} - \phi_v \psi_{uv}}{\Delta} + \sigma,$$

so that the second and third equations of (3.5) can be written

$$(3.9) \quad B = -2(\log \Delta)_u + 3\rho, \quad C = 2(\log \Delta)_v - 3\sigma.$$

The partial differential equations (3.6) are easily integrated, and the result may be written in the form

$$(3.10) \quad U_1\phi + U_2\psi = 1, \quad V_1\phi + V_2\psi = -1;$$

therefore  $\phi, \psi$  must have the forms

$$(3.11) \quad \phi = \frac{U_2 + V_2}{U_1V_2 - U_2V_1}, \quad \psi = -\frac{U_1 + V_1}{U_1V_2 - U_2V_1},$$

where  $U_1V_2 - U_2V_1 \neq 0$ ; that is,  $U_1/U_2$  and  $V_1/V_2$  are not equal to the same constant, nor do we have  $U_1=0$  and  $U_2=0$  or  $V_1=0$  and  $V_2=0$ .

From (3.11) we calculate

$$(3.12) \quad \begin{aligned} \phi_u &= V_2 \frac{\begin{vmatrix} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{vmatrix}}{\begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}^2}, & \phi_v &= -U_2 \frac{\begin{vmatrix} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{vmatrix}}{\begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}^2}, \\ \psi_u &= -V_1 \frac{\begin{vmatrix} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{vmatrix}}{\begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}^2}, & \psi_v &= U_1 \frac{\begin{vmatrix} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{vmatrix}}{\begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}^2}, \end{aligned}$$

therefore, by the last equation of (3.5),

$$(3.13) \quad \Delta = \frac{\begin{vmatrix} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{vmatrix} \begin{vmatrix} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{vmatrix}}{\begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}^3}.$$

By substitution of (3.12) in (3.8),

$$(3.14) \quad \begin{aligned} \rho &= \frac{\partial}{\partial u} \log \begin{vmatrix} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{vmatrix} - 2 \frac{\partial}{\partial u} \log \begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}, \\ \sigma &= \frac{\partial}{\partial v} \log \begin{vmatrix} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{vmatrix} - 2 \frac{\partial}{\partial v} \log \begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}. \end{aligned}$$

Substituting (3.13), (3.14) in (3.9), we find

$$(3.15) \quad \begin{aligned} B &= \frac{\partial}{\partial u} \log \left| \begin{array}{cc} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{array} \right| - 2 \frac{\partial}{\partial u} \log \left| \begin{array}{cc} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{array} \right|, \\ C &= 2 \frac{\partial}{\partial v} \log \left| \begin{array}{cc} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{array} \right| - \frac{\partial}{\partial v} \log \left| \begin{array}{cc} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{array} \right|, \end{aligned}$$

that is, as abbreviated by the notation (1.5),

$$(3.16) \quad B = \frac{\partial}{\partial u} (\log I - 2 \log II), \quad C = \frac{\partial}{\partial v} (2 \log I - \log II).$$

We thus have the result:

*In order that a velocity family expressed in minimal coordinates be linear, it is necessary and sufficient that the coefficients  $B, C$  in (2.13) have the special form (3.15) or (3.16).*

To complete the imposition of property (3), in addition to the property (2) of linearity, we must particularize our velocity family by the additional condition (2.17):  $C_u - B_v = 2kF$ . Substituting from (3.16), this gives the result stated in formula (1.7):

$$(3.17) \quad 2kF = \frac{\partial^2}{\partial u \partial v} (\log I + \log II).$$

We find by direct calculation:

$$(3.18) \quad \begin{aligned} \frac{\partial^2}{\partial u \partial v} \log I &= \frac{\left| \begin{array}{cc} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{array} \right| \cdot \left| \begin{array}{cc} V'_1 & U''_1 \\ V'_2 & U''_2 \end{array} \right|}{\left| \begin{array}{cc} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{array} \right|^2} \\ &+ \frac{\left| \begin{array}{cc} U_1 + V_1 & U''_1 \\ U_2 + V_2 & U''_2 \end{array} \right| \cdot \left| \begin{array}{cc} U'_1 & V'_1 \\ U'_2 & V'_2 \end{array} \right|}{\left| \begin{array}{cc} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{array} \right|^2}. \end{aligned}$$

The determinants which appear in the numerators are among the six in the matrix

$$\left\| \begin{array}{cccc} U_1 + V_1 & U'_1 & V'_1 & U''_1 \\ U_2 + V_2 & U'_2 & V'_2 & U''_2 \end{array} \right\|,$$

which obey the well known identity<sup>(13)</sup>:

<sup>(13)</sup> The same as the one which governs Plücker line coordinates, being obeyed by the six determinants of second order in any two-by-four matrix.

$$(3.19) \quad \begin{vmatrix} U_1 + V_1 & U'_1 \\ U_2 + V_2 & U'_2 \end{vmatrix} \cdot \begin{vmatrix} V'_1 & U''_1 \\ V'_2 & U''_2 \end{vmatrix} + \begin{vmatrix} U_1 + V_1 & U''_1 \\ U_2 + V_2 & U''_2 \end{vmatrix} \cdot \begin{vmatrix} U'_1 & V'_1 \\ U'_2 & V'_2 \end{vmatrix} \\ = \begin{vmatrix} U_1 + V_1 & V'_1 \\ U_2 + V_2 & V'_2 \end{vmatrix} \cdot \begin{vmatrix} U'_1 & U''_1 \\ U'_2 & U''_2 \end{vmatrix}.$$

Therefore

$$(3.20) \quad \frac{\partial^2}{\partial u \partial v} \log I = \frac{II}{I^2} \begin{vmatrix} U'_1 & U''_1 \\ U'_2 & U''_2 \end{vmatrix}.$$

Similarly we find

$$(3.21) \quad \frac{\partial^2}{\partial u \partial v} \log II = \frac{I}{II^2} \begin{vmatrix} V'_1 & V''_1 \\ V'_2 & V''_2 \end{vmatrix}.$$

These are the formulas stated as (1.8); substituted in (3.17), they give the expanded form (1.7') for  $2kF$ .

The proof of our main results is thus completed.

4. **General coordinates.** It is interesting to see how our formulas look in general coordinates  $u, v$ , instead of minimal coordinates.

Using the general formula<sup>(14)</sup> for geodesic curvature  $1/\rho$ , the condition (2.10) for a velocity family gives the following characteristic differential equation for such a family:

$$(4.1) \quad \begin{aligned} W^2 v'' &= W(P + Qv')(E + 2Fv' + Gv'^2) \\ &+ (F + Gv') \left[ \frac{1}{2} E_u + E_v v' + (F_v - \frac{1}{2} G_u) v'^2 \right] \\ &- (E + Fv') \left[ (F_u - \frac{1}{2} E_v) + G_u v' + \frac{1}{2} G_v v'^2 \right]. \end{aligned}$$

The imposition of property (3) is completed by (2.11), where  $K$  is to be thought of as expressed in terms of  $E, F, G$  and their first and second partial derivatives<sup>(15)</sup>:

$$(4.2) \quad Q_u - P_v = (k - K)W.$$

We have also to express property (2), of linearity, in general coordinates. It is a known result<sup>(16)</sup> that the most general linear family has the form

$$(4.3) \quad v'' = A + Bv' + Cv'^2 + Dv'^3,$$

where  $A, B, C, D$  are functions of  $u, v$  which obey the conditions

$$(4.4) \quad \begin{aligned} (AC - A_v)_v - (AD + \frac{1}{3}C_u - \frac{2}{3}B_v)_u \\ + B(AD + \frac{1}{3}C_u - \frac{2}{3}B_v) - A(BD + D_u) = 0, \end{aligned}$$

<sup>(14)</sup> Blaschke, loc. cit., p. 117.

<sup>(15)</sup> See formula (4.11).

<sup>(16)</sup> Due to Lie and R. Liouville. See E. Kasner [1].

$$(4.5) \quad - (BD + D_u)_u + (AD + \frac{2}{3}C_u - \frac{1}{3}B_v)_v \\ + C(AD + \frac{2}{3}C_u - \frac{1}{3}B_v) - D(AC - A_v) = 0.$$

The formula (4.1) is of the type (4.3) with

$$(4.6) \quad \begin{aligned} A &= \{ WPE + \frac{1}{2}FE_u - E(F_u - \frac{1}{2}E_v) \} / W^2, \\ B &= \{ 2WPF + WQE + FE_v + \frac{1}{2}GE_u - F(F_u - \frac{1}{2}E_v) - EG_u \} / W^2, \\ C &= \{ WPG + 2WQF + GE_v + F(F_v - \frac{1}{2}G_u) - \frac{1}{2}EG_v - FG_u \} / W^2, \\ D &= \{ WQG + G(F_v - \frac{1}{2}G_u) - \frac{1}{2}FG_v \} / W^2. \end{aligned}$$

The conditions (4.4), (4.5) of linearity give two partial differential equations involving  $E, F, G, P, Q$ .

In summary: configurations  $(\mathcal{Y}, S)$  having properties (2) and (3) are characterized in general coordinates by

$$S: ds^2 = Edu^2 + 2Fdudv + dvG^2,$$

$$\mathcal{Y}: \text{ of type (4.1),}$$

where the five functions  $E, F, G, P, Q$  obey the three partial differential equations (4.2), (4.4), (4.5), in which  $A, B, C, D$  are defined by (4.6). These equations are of the third order in  $E, F, G$  and the second order in  $P, Q$ .

Of course, the general solution is obtainable by applying an arbitrary transformation  $u_1 = \phi(u, v)$ ,  $v_1 = \psi(u, v)$  to the formulas based on minimal coordinates.

We may inquire also as to the form our equations have in the particular coordinate system  $(u, v)$  where the equations of  $\mathcal{Y}$  are linear<sup>(17)</sup>:  $v = au + b$ , or  $v'' = 0$ . In (4.6), we have then to write

$$(4.7) \quad A = 0, \quad B = 0, \quad C = 0, \quad D = 0,$$

which, in addition to (4.2), give a system of five conditions on  $E, F, G, P, Q$ . From these we easily eliminate  $P, Q$  and obtain the following three conditions on  $E, F, G$ :

$$(4.8) \quad (\frac{1}{2}EG - F^2) \frac{E_u}{E} + \frac{1}{2}FE_v + FF_u - EF_v - \frac{1}{2}EG_u + \frac{1}{2}EF \frac{G_v}{G} = 0,$$

$$(4.9) \quad (\frac{1}{2}EG - F^2) \frac{G_v}{G} + \frac{1}{2}FG_u + FF_v - GF_u - \frac{1}{2}GE_v + \frac{1}{2}FG \frac{E_u}{E} = 0,$$

$$(4.10) \quad 2W^3 \frac{\partial}{\partial v} \frac{FE_u - EF_u}{WE} + 2W^3 \frac{\partial}{\partial u} \frac{FG_v - GF_v}{WG} - \begin{vmatrix} E & E_u & E_v \\ F & F_u & F_v \\ G & G_u & G_v \end{vmatrix} = 4kW^4.$$

<sup>(17)</sup> Of course, this coordinate system is determined only up to an arbitrary projective transformation of  $u, v$ .

In the derivation of (4.10), we use the formula of G. Frobenius for Gaussian curvature<sup>(18)</sup>:

$$(4.11) \quad K = -\frac{1}{4W^4} \begin{vmatrix} E & E_u & E_v \\ F & F_u & F_v \\ G & G_u & G_v \end{vmatrix} - \frac{1}{2W} \left\{ \frac{\partial}{\partial v} \frac{E_v - F_u}{W} - \frac{\partial}{\partial u} \frac{F_v - G_u}{W} \right\}.$$

In summary, (4.8)–(4.10) are necessary and sufficient conditions on the  $E, F, G$  of a surface in order that the curve family defined by the general linear equation,  $v = au + b$ , have the property  $\mathcal{E} = k\mathcal{A}$ .

(4.8) and (4.9) are necessary and sufficient in order that  $v = au + b$  shall be a velocity family.

**5. Geometric construction for the property  $\mathcal{E} = k\mathcal{A}$ .** In the case  $k = 0$ , it is well known that the curve families  $\mathcal{F}$  which have the property:  $\mathcal{E} = 0$  or  $A + B + C = \pi$  for every triangle  $ABC$  of  $\mathcal{F}$ , are exactly the *isogonal* families<sup>(19)</sup>. These consist of the totality of  $\infty^2$  trajectories under every possible constant angle  $\theta$  of any given family  $\alpha$  of  $\infty^1$  curves.

It is easy to generalize this construction to the case  $k \neq 0$ . Take any net of curves upon an arbitrary surface  $S$ , composed of two families  $\alpha, \beta$  of  $\infty^1$  curves. Construct a trajectory  $T$  of the family  $\alpha$ , not under constant angle, but rather so that the angle  $\theta$  between  $T$  and  $\alpha$  decreases by  $k$  times the element of area swept out by the arc of the curve  $\beta_p$  of  $\beta$  which passes through the point  $p$  describing  $T$  and extends to the intersection  $m$  of  $\beta_p$  with any chosen fixed curve  $\alpha_0$  of  $\alpha$ .

That is—with reference to a figure easily drawn by the reader—we have, in integrated form, the law

$$(5.1) \quad \theta_2 - \theta_1 = -k \cdot \text{area } p_1 p_2 m_1 m_2$$

for the construction of  $T$ . Evidently, this law determines the formation of  $T$ , element by element, when any initial point and direction are given; therefore the totality of trajectories  $T$  is a family  $\mathcal{T}$  of  $\infty^2$  curves. It is very easy to give by means of a figure a proof of the property:  $\mathcal{E} = k\mathcal{A}$  for every triangle of  $\mathcal{T}$ .

It remains to be shown that every family  $\mathcal{T}$  with the property  $\mathcal{E} = k\mathcal{A}$  is obtainable by a construction of the type just described. This is readily done by the following converse reasoning. Construct the pencil  $\Pi$  of curves of  $\mathcal{T}$  through a fixed point  $p$ . In the region  $R$  covered by  $\Pi$ , construct any family  $\beta$  of  $\infty^1$  curves all of which intersect a fixed base curve  $\alpha_0$ . At each point of  $R$  construct a direction  $\delta$  according to the following law: (i) at the points of  $\alpha_0$ ,  $\delta$  shall coincide with the tangential direction to  $\alpha_0$ ; (ii) the angle  $\theta$  between  $\Pi$  and  $\delta$  shall vary according to the law (5.1). We thus have a field of directions

<sup>(18)</sup> Blaschke, loc. cit., p. 79.

<sup>(19)</sup> G. Scheffers, *Isogonalkurven, Äquitangentialkurven und komplexe Zahlen*, Mathematische Annalen, vol. 60 (1905), p. 504.



$\delta$ , whose integration gives a family of  $\infty^1$  curves  $\alpha$ , including  $\alpha_0$ . If now  $T$  is any curve of the family  $\mathcal{F}$  lying in the region  $R$ , it is evident, by applying the property  $\mathcal{E} = kA$  to the triangle formed by any arc  $p_1p_2$  of  $T$  and the curves  $pp_1, pp_2$  of  $\Pi$ , that  $T$  traverses  $\alpha$  according to the law (5.1).

Thus the law (5.1) holds as long as  $T$  lies in the region  $R$  covered by  $\Pi$ . We can extend this region by applying the same reasoning to the pencil of curves of  $\mathcal{F}$  which pass through any other fixed point  $p'$  of  $R$ , and repeating this procedure any finite number of times.

6. **Higher dimensions,  $n > 2$ .** We conclude with a statement of the analogous problem for higher dimensional spaces, which we hope to consider in a future paper.

Let  $\mathcal{F}$  denote a linear family of curves in a space of  $n > 2$  dimensions; that is, let  $\mathcal{F}$  be depictable as the totality of straight lines of a flat projective  $n$ -space  $P$ . It is required to impose on the space  $P$  a Riemannian metric  $R: ds^2 = g_{ij}dx^i dx^j$ , so that, in every triangle of  $P$ ,  $\mathcal{E} = kA$  with  $k$  a preassigned nonzero constant, angles and areas being measured according to  $R$ .

Certainly, a sufficient condition is that the space  $R$  have constant Riemannian curvature and that  $\mathcal{F}$  consist of its geodesics. In other words,  $R$  shall be the Cayley metric based on any fixed quadric  $Q$ :  $\text{dist } ab = \{2(-k)^{1/2}\}^{-1} \cdot \log(abp_1p_2)$ , where  $p_1, p_2$  are the intersection points of the line  $ab$  with  $Q$ , and the parenthesis denotes an anharmonic ratio. For the Cayley metric is a typical one of constant curvature, and the straight lines are its geodesics.

It remains to be seen whether, for  $n > 2$ , the Cayley metric is the most general one which can be imposed on the projective space  $P$  so that  $\mathcal{E} = kA$ . We reserve this problem for future consideration.

It may be remarked that for  $k = 0$  and  $n > 2$  the property  $\mathcal{E} = 0$ —that is, the property that the angle-sum of every triangle of  $\mathcal{F}$  is two right angles—implies that  $\mathcal{F}$  is linear<sup>(20)</sup>. The author has proved that, furthermore, it must be possible to represent the Riemann space  $R$  conformally on a euclidean space  $E$  so that  $\mathcal{F}$  corresponds to the straight lines of  $E$ <sup>(20)</sup>.

#### REFERENCES

1. E. Kasner, *A characteristic property of isothermal systems of curves*, Mathematische Annalen, vol. 59 (1904), pp. 352–354.
2. J. Douglas, *A criterion for the conformal equivalence of a Riemann space to a euclidean space*, these Transactions, vol. 27 (1925), pp. 299–306.
3. J. Radon, *Kurvennetze auf Flächen und im Raume von Riemann*, Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität, vol. 5 (1927), pp. 45–53.
4. C. H. Rowe, *On certain systems of curves in Riemannian space*, Journal de Mathématiques Pures et Appliquées, vol. 12 (1933), pp. 283–308.

BROOKLYN, N. Y.

<sup>(20)</sup> Reference [2].

# ON STRONG SUMMABILITY OF FOURIER SERIES

BY  
OTTO SZÁSZ

1. A series  $\sum_0^\infty A_n$ , or the corresponding sequence of partial sums  $s_n = \sum_0^n A_n$ ,  $n=0, 1, 2, \dots$ , is said to be strongly summable  $(C, 1)$  with index  $k$  to the sum  $s$  if  $k > 0$  and

$$(1.1) \quad \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_0^n |s_\nu - s|^k = 0 \quad (1).$$

It follows from Hölder's inequality that the larger  $k$  the stronger is the assertion (1.1). Furthermore, for  $k=1$ , (1.1) evidently implies  $(C, 1)$  summability to the sum  $s$ .

Suppose now that  $f(t)$  is a periodic function of the class  $L$ . Let its Fourier series be

$$(1.2) \quad f(t) \sim \frac{1}{2}a_0 + \sum_1^\infty (a_\nu \cos \nu t + b_\nu \sin \nu t) = \sum_0^\infty A_\nu(t);$$

let

$$(1.3) \quad \phi(x, t) = \frac{1}{2} \{ f(x+t) + f(x-t) - 2s \}.$$

Hardy and Littlewood proved (1913):

**THEOREM I.** *The Fourier series of an integrable function  $f(t)$  is strongly summable  $(C, 1)$  with index 2 at a point  $x$  if  $f(t)$  is of integrable square in a neighborhood of  $x$  and if for some  $s$*

$$(1.4) \quad \int_0^t \{ \phi(x, u) \}^2 du = o(t) \quad \text{as } t \downarrow 0.$$

In this paper we shall restrict ourselves to the index  $k=2$ , and speak simply of this case as "strong summability." For generalizations of Theorem I and for further references consult Hardy and Littlewood [2] and Zygmund [5, chap. 10].

For the special case in which  $\phi(t) \rightarrow 0$  as  $t \downarrow 0$ , Fejér [1] recently gave two new proofs of the strong summability of the series (1.2) at  $t=x$ . We shall simplify his device and use it to give two new proofs of Theorem I. The essence of Fejér's method is to introduce double integrals with positive kernels while using the  $(C, 3)$  and Abel summability methods. Replacing the partial sums  $s_n$  by  $s_n - \frac{1}{2}A_n$  we get simpler (and also positive) kernels.

Presented to the Society, December 26, 1939; received by the editors February 17, 1940.

(1) For generalizations to other summability methods cf. [3, §§7, 8 and 11]. Numbers in brackets refer to the bibliography at the end of the paper.

In the last section we prove still another theorem of Hardy and Littlewood [2]:

THEOREM II. *If*

$$(1.5) \quad \int_0^t |\phi(x, u)| du = o(t) \quad \text{as } t \downarrow 0,$$

then

$$(1.6) \quad \sum_0^n (s_\nu(x) - s)^2 = o(n \log n) \quad \text{as } n \rightarrow \infty.$$

Our proof is shorter and simpler, not involving complex function theory. Hardy and Littlewood also proved, by constructing examples, that (1.6) is the sharpest asymptotic estimate implied by the assumption (1.5).

2. We prove first the following lemma.

LEMMA. *Let  $s_0^* = 0$ ,  $s_n^* = s_n - \frac{1}{2}A_n$ ,  $n = 1, 2, \dots$ ; if  $\lim_{n \rightarrow \infty} A_n = 0$ , and if one of the sequences  $s_n$ ,  $s_n^*$  is strongly summable, so is the other.*

This follows from the identities

$$\begin{aligned} \sum_1^n (s_\nu - s)^2 - \sum_1^n (s_\nu^* - s)^2 &= \frac{1}{2} \sum_1^n A_\nu (s_\nu + s_\nu^* - 2s) \\ &= \sum_1^n A_\nu (s_\nu - s) - \frac{1}{4} \sum_1^n A_\nu^2 \\ &= \sum_1^n A_\nu (s_\nu^* - s) + \frac{1}{4} \sum_1^n A_\nu^2. \end{aligned}$$

In view of this lemma, we may deal with  $s_n^*(x)$  instead of  $s_n(x)$  while discussing the series (1.2). Now

$$(2.1) \quad s_n^*(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x+t) \cot \frac{1}{2}t \sin nt \, dt = \frac{1}{\pi} \int_0^{\pi} \psi(x, t) \cot \frac{1}{2}t \sin nt \, dt,$$

where  $\psi(x, t) = \frac{1}{2} \{f(x+t) - f(x-t)\} = \psi(t)$ . Hence

$$s_n^*(x)^2 = \frac{1}{\pi^2} \int_0^{\pi} \int_0^{\pi} \psi(t)\psi(u) \cot \frac{1}{2}t \cot \frac{1}{2}u \sin nt \sin nu \, dt du,$$

and

$$(2.2) \quad \sum_1^n (n+1-\nu)^2 s_\nu^*(x)^2 = \frac{1}{\pi^2} \int_0^{\pi} \int_0^{\pi} \psi(t)\psi(u) \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) \, dt du,$$

where  $R_n(t, u) = \sum_{\nu=1}^n (n+1-\nu)^2 \sin \nu t \sin \nu u$ .

If  $f(t) \equiv 1$ , then  $A_n(t) \equiv 1$ ,  $\nu = 1, 2, \dots$ , and  $s_n^*(x) = 1$ ,  $\nu = 1, 2, \dots$ ; hence from (2.2)

$$(2.3) \quad \sum_1^n (n+1-\nu)^2 = \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) dt du \\ = \frac{1}{6}n(n+1)(2n+1).$$

Now

$$(2.4) \quad R_n(t, u) > 0 \quad \text{for } 0 < t < \pi, 0 < u < \pi;$$

the proof is elementary (cf. [4, §2]).

As a first application of (2.2), (2.3) and (2.4) we get:

$$\text{If } |f(t)| \leq 1 \text{ in } |t| \leq \pi, \text{ then } \sum_1^n (n+1-\nu)^2 s_n^*(x)^2 \leq \sum_1^n \nu^2.$$

Also from (2.1)

$$1 = \frac{1}{\pi} \int_0^\pi \cot \frac{1}{2}t \sin nt dt, \quad n = 1, 2, \dots;$$

hence

$$(2.5) \quad s_n^* - s = \frac{1}{\pi} \int_0^\pi \phi(x, t) \cot \frac{1}{2}t \sin nt dt, \quad n = 1, 2, \dots,$$

and, writing  $\psi(t)$  for  $\psi(x, t)$ ,

$$\sum_1^n (n+1-\nu)^2 (s_n^* - s)^2 = \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \phi(t)\phi(u) \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) dt du \\ \equiv I_n(\phi).$$

Now (2.4) yields

$$(2.6) \quad |I_n(\phi)| \leq I_n(\bar{\phi}),$$

whenever  $|\phi(t)| \leq \bar{\phi}(t)$  in  $0 < t < \pi$ .

The proof of strong summability at a point where  $\phi(x, t) \rightarrow 0$  as  $t \downarrow 0$  now follows as in Fejér's method. We note first that  $I_n(\phi) = o(n^2)$  as  $n \rightarrow \infty$  is a necessary and sufficient condition for the strong summability of the series (1.2) at  $t=x$ , or, what is the same, of the cosine series of  $\phi(t)$  at  $t=0$ . This follows from the following general inequalities for an arbitrary sequence of positive quantities  $p_r \geq 0$ :

$$(2.7) \quad (n+1)^{-2} \sum_0^n (n+1-\nu)^2 p_\nu \leq \sum_0^n p_\nu \leq n^{-2} \sum_0^{2n} (2n-\nu)^2 p_\nu.$$

Next, (2.6) yields the following theorem:

Whenever  $|\phi(t)| \leq \bar{\phi}(t)$  in  $0 < t < \pi$ , the strong summability of the cosine series of  $\phi(t)$  at  $t=0$  implies that of the series of  $\bar{\phi}(t)$ .

If now  $\phi(t) \rightarrow 0$  as  $t \downarrow 0$ , then there is an interval  $0 \leq t \leq \delta$  in which  $\phi(t)$  is bounded. We choose the majorant function  $\bar{\phi}$  to be

$$\bar{\phi}(t) = \begin{cases} \max_{0 \leq \tau \leq t} |\phi(\tau)| & \text{if } 0 \leq t \leq \delta \\ |\phi(t)| & \text{if } \delta < t \leq \pi. \end{cases}$$

Now  $\bar{\phi}$  is continuous at  $t=0$  and monotonic in  $0 < t < \delta$ ; hence its cosine series converges at  $t=0$ , and it is, consequently, strongly summable. Thus the series (1.2) is strongly summable at  $t=x$ .

3. The symmetry of the integrand gives

$$I_n(\phi) = \frac{1}{\pi^2} \iint_{0 \leq t \leq u \leq \pi} \cdots + \frac{1}{\pi^2} \iint_{0 \leq u \leq t \leq \pi} \cdots = \frac{2}{\pi^2} \iint_{0 \leq t \leq u \leq \pi} \cdots$$

Furthermore for the function

$$(3.1) \quad \phi_1(t) = \begin{cases} 0 & \text{for } 0 < t < \delta \\ \phi(t) & \text{for } \delta < t < \pi, \end{cases}$$

$$I_n(\phi_1) = \frac{2}{\pi^2} \iint_{\delta \leq t \leq u \leq \pi} \phi(t)\phi(u) \cdots = o(n^3) \quad \text{as } n \rightarrow \infty,$$

since the cosine series of  $\phi_1(t)$  converges to 0 at  $t=0$ . This yields the following result:

*A necessary and sufficient condition that (1.2) be strongly summable at  $t=x$  is that for a fixed  $\delta > 0$*

$$I_n^{(\delta)}(\phi) = \frac{1}{\pi^2} \iint_{0 \leq t \leq u \leq \delta} \phi(t)\phi(u) \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) dt du = o(n^3)$$

as  $n \rightarrow \infty$ .

We now use this criterion to prove Theorem I.

Schwarz's inequality yields

$$I_n^{(\delta)}(\phi)^2 \leq \frac{1}{\pi^4} \iint_{0 \leq t \leq u \leq \delta} \phi(t)^2 \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) dt du \\ \cdot \iint_{0 \leq t \leq u \leq \delta} \phi(u)^2 \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) dt du.$$

Hence

$$\begin{aligned}
 I_n^{(1)}(\phi) &\leq \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \phi(t)^2 \cot \frac{1}{2}t \cot \frac{1}{2}u R_n(t, u) dt du \\
 &\leq \frac{1}{\pi^2} \int_0^\pi \left\{ \phi(t)^2 \cot \frac{1}{2}t \int_0^\pi \cot \frac{1}{2}u R_n(t, u) du \right\} dt.
 \end{aligned}$$

But

$$\begin{aligned}
 \int_0^\pi \cot \frac{1}{2}u R_n(t, u) du &= \sum_1^n (n+1-v)^2 \sin vt \int_0^\pi \cot \frac{1}{2}u \sin vu du \\
 &= \pi \sum_1^n (n+1-v)^2 \sin vt;
 \end{aligned}$$

and the relation

$$\int_0^\pi \phi(t)^2 \cot \frac{1}{2}t \sum_1^n (n+1-v)^2 \sin vt dt = o(n^2)$$

follows from Lebesgue's theorem on  $(C, 1)$  summability applied to  $\phi(t)^2$ , using (1.4) and (2.7). This proves Theorem I.

4. We shall now apply the Abel-Poisson summability method. From (2.1) for  $0 < r < 1$

$$\begin{aligned}
 \sum_1^\infty (s_n^*(x))^2 r^n &= \frac{1}{\pi^2} \int_0^\pi \int_0^\pi \psi(t)\psi(u) \cot \frac{1}{2}t \cot \frac{1}{2}u \left( \sum_0^\infty \sin nt \sin nu r^n \right) dt du \\
 &= \frac{4r(1-r^2)}{\pi^2} \int_0^\pi \int_0^\pi \psi(t)\psi(u) \cos^2 \frac{1}{2}t \cos^2 \frac{1}{2}u \\
 &\quad \cdot [1 - 2r \cos(u-t) + r^2]^{-1} [1 - 2r \cos(u+t) + r^2]^{-1} dt du.
 \end{aligned}$$

Putting  $f(t) \equiv 1$ , we get

$$\begin{aligned}
 \frac{r}{(1-r)} &= \frac{4r(1-r^2)}{\pi^2} \int_0^\pi \int_0^\pi \cos^2 \frac{1}{2}t \cos^2 \frac{1}{2}u [1 - 2r \cos(u-t) + r^2]^{-1} \\
 &\quad \cdot [1 - 2r \cos(u+t) + r^2]^{-1} dt du.
 \end{aligned}$$

The integrand will be denoted by  $P(t, u; r)$ . Evidently

$$(4.1) \quad P(t, u; r) > 0 \quad \text{for } 0 < t < \pi, 0 < u < \pi, 0 < r < 1.$$

If  $|f(t)| \leq 1$ ,  $0 < t < 2\pi$ , this yields

$$(4.2) \quad \sum_1^\infty s_n^*(x)^2 r^n \leq \sum_1^\infty r^n \quad \text{for } 0 < r < 1.$$

Similarly from (2.5)



$$\sum_1^\infty (s_n^* - s)^2 r^n = \frac{4r(1-r^2)}{\pi^2} \int_0^\pi \int_0^\pi \phi(t)\phi(u)P(t, u; r)dtdu \equiv A(\phi; r),$$

and (4.1) gives  $A(\phi; r) \leq A(\bar{\phi}; r)$  whenever  $|\phi(t)| \leq \bar{\phi}(t)$ . We first remark that

$$A(\phi; r) = o\left(\frac{1}{1-r}\right) \quad \text{as } r \uparrow 1$$

is a necessary and sufficient condition for strong summability; i.e., for  $\sum_0^n (s_n^* - s)^2$  to be  $o(n)$  as  $n \rightarrow \infty$ .

The necessity is obvious; the sufficiency follows from the inequality (valid for any  $p, \geq 0$ ):

$$\sum_0^n p_n \leq \left(1 - \frac{1}{n}\right)^{-n} \sum_0^n p_n \left(1 - \frac{1}{n}\right)^n \leq 4 \sum_0^\infty p_n \left(1 - \frac{1}{n}\right)^n, \quad n \geq 2.$$

If now  $\phi(t) \rightarrow 0$  as  $t \downarrow 0$ , then, using the same majorant as in §2, we obtain still another proof of strong summability at points where  $\phi(t) \rightarrow 0$ . It is similar to Fejér's second proof except that we use a simpler kernel.

To prove Theorem I we observe that for the function  $\phi_1(t)$  of (3.1) evidently

$$(4.3) \quad A(\phi_1; r) = \frac{4r(1-r^2)}{\pi^2} \int_\delta^\pi \int_\delta^\pi \phi(t)\phi(u)P(t, u; r)dtdu = o\left(\frac{1}{1-r}\right).$$

This together with the symmetry of the integrand gives (as in §3):

*A necessary and sufficient condition for strong summability of (1.2) at  $t=x$  is that for a fixed  $\delta > 0$*

$$A_\delta(\phi; r) \equiv \frac{4r(1-r^2)}{\pi^2} \int_0^\delta \int_0^\delta \phi(t)\phi(u)P(t, u; r)dtdu = o\left(\frac{1}{1-r}\right).$$

Again using Schwarz's inequality, we obtain

$$(4.4) \quad \begin{aligned} |A_\delta(\phi; r)| &\leq \frac{4r(1-r^2)}{\pi^2} \int_0^\delta \int_0^\delta \phi(t)^2 P(t, u; r)dtdu \\ &\leq \frac{4r(1-r^2)}{\pi^2} \int_0^\delta \left[ \phi(t)^2 \int_0^\pi P(t, u; r)du \right] dt. \end{aligned}$$

But

$$\begin{aligned} 4r(1-r^2) \int_0^\pi P(t, u; r)du &= \cot \frac{1}{2}t \int_0^\pi \cot \frac{1}{2}u \left( \sum_1^\infty \sin nt \sin nu r^n \right) du \\ &= \pi \cot \frac{1}{2}t \sum_1^\infty \sin nt r^n. \end{aligned}$$

The right-hand side of (4.4) is  $o(1-r)^{-1}$  since the cosine series of  $\phi(t)^2$  is Poisson summable at  $t=0$  under assumption (1.4). We have thus proved Theorem I once again.

5. In this section we assume

$$(5.1) \quad \int_0^t |\phi(x, u)| du \equiv \Phi(x, t) = o(t) \quad \text{as } t \downarrow 0.$$

From (4.3)

$$(5.2) \quad A(\phi; r) \leq A(|\phi|; r) = \frac{4r(1-r^2)}{\pi^2} \int_0^\delta \int_0^\delta |\phi(t)\phi(u)| P(t, u; r) dt du + o\left(\frac{1}{1-r}\right).$$

The first term on the right is

$$(5.3) \quad \begin{aligned} A_1(|\phi|; r) &= \frac{4r(1-r^2)}{\pi^2} \int_0^\delta \int_0^\delta |\phi(t)\phi(u)| \\ &\quad \frac{\cos^2 \frac{1}{2}t \cos^2 \frac{1}{2}u dt du}{[(1-r)^2 + 4r \sin^2 \frac{1}{2}(u-t)][(1-r)^2 + 4r \sin^2 \frac{1}{2}(u+t)]} \\ &\leq \frac{4r(1-r^2)}{\pi^2} \int_0^\delta \int_0^\delta |\phi(t)\phi(u)| \\ &\quad \cdot \left[(1-r)^2 + \frac{r}{\pi^2}(u-t)^2\right]^{-1} \left[(1-r)^2 + \frac{r}{\pi^2}(u+t)^2\right]^{-1} dt du \end{aligned}$$

assuming  $0 < \delta < \pi/2$ . Let  $1-r < \delta$ , and decompose the range of integration into  $0 \leq u+t \leq 1-r$  and  $1-r \leq u+t \leq 2\delta$ . For the first part, using (5.1), we get

$$(5.4) \quad \iint_{0 \leq u+t \leq 1-r} \dots < (1-r)^{-4} \left( \int_0^{1-r} |\phi(t)| dt \right)^2 = o((1-r)^{-2}).$$

Hence, for  $r \uparrow 1$ ,

$$(5.5) \quad \begin{aligned} A_1(|\phi|; r) &< o\left(\frac{1}{1-r}\right) + 4(1-r^2) \iint_{1-r \leq u+t \leq 2\delta} |\phi(t)\phi(u)| (u+t)^{-2} \\ &\quad \cdot [(1-r)^2 + r(u-t)^2]^{-1} dt du. \end{aligned}$$

Now, the last integral is

$$(5.6) \quad \begin{aligned} 2 \iint_{0 \leq t \leq u}^{1-r \leq u+t \leq 2\delta} \dots &\leq 2 \iint_{0 \leq t \leq u} |\phi(t)\phi(u)| u^{-2} \\ &\quad \cdot [(1-r)^2 + r(u-t)^2]^{-1} dt du \equiv 2B_\delta(r). \end{aligned}$$

Furthermore

$$\begin{aligned} B_\delta(r) &\leq \int_{(1-r)/2}^{2\delta} u^{-2} |\phi(u)| \left( \int_0^u |\phi(t)| [(1-r)^2 + r(u-t)^2]^{-1} dt \right) du \\ &< (1-r)^{-2} \int_{(1-r)/2}^{2\delta} u^{-2} |\phi(u)| \Phi(u) du \\ &= (1-r)^{-2} O \left( \int_{(1-r)/2}^{2\delta} u^{-1} |\phi(u)| du \right), \end{aligned}$$

and

$$\begin{aligned} \int_{(1-r)/2}^{2\delta} u^{-1} |\phi(u)| du &= u^{-1} \Phi(u) \Big|_{(1-r)/2}^{2\delta} + \int_{(1-r)/2}^{2\delta} u^{-2} \Phi(u) du \\ &= O(1) + \int_{(1-r)/2}^{2\delta} u^{-2} \Phi(u) du. \end{aligned}$$

Thus from (5.2), (5.5) and (5.6), as  $r \uparrow 1$

$$\begin{aligned} A(\phi; r) &< o \left( \frac{1}{1-r} \right) + o \left( \frac{1}{1-r} \right) + O \left( \frac{1}{1-r} \right) \\ &\quad + \frac{1}{1-r} O \left( \int_{(1-r)/2}^{2\delta} u^{-2} \Phi(u) du \right). \end{aligned}$$

Finally

$$\int_{(1-r)/2}^{2\delta} u^{-2} \Phi(u) du = \left( \int_{(1-r)/2}^{\epsilon(r)} + \int_{\epsilon(r)}^{2\delta} \right) u^{-2} \Phi(u) du \equiv C_1(r) + C_2(r),$$

where we may assume

$$\frac{1}{2}(1-r) < \epsilon(r) \equiv \exp [-(\log(1-r)^{-1})^{1/2}] < 2\delta.$$

Now

$$\begin{aligned} C_1(r) &\leq \max_{u \leq \epsilon(r)} u^{-1} \Phi(u) \int_{(1-r)/2}^{\epsilon(r)} u^{-1} du = \max_{u \leq \epsilon(r)} u^{-1} \Phi(u) [\log \epsilon(r) - \log \frac{1}{2}(1-r)] \\ &= \max_{u \leq \epsilon(r)} u^{-1} \Phi(u) \left[ \log \frac{2}{1-r} - \left( \log \frac{1}{1-r} \right)^{1/2} \right] \\ &= o \left( \log \frac{1}{1-r} \right) \quad \text{as } r \uparrow 1, \end{aligned}$$

and

$$C_2(r) = O \left( \int_{\epsilon(r)}^{2\delta} u^{-1} du \right) = O \left( \log 2\delta + \left( \log \frac{1}{1-r} \right)^{1/2} \right) = o \left( \log \frac{1}{1-r} \right)$$

as  $r \uparrow 1$ . Summarizing,

$$A(\phi; r) < O\left(\frac{1}{1-r}\right) + o\left(\frac{1}{1-r} \log \frac{1}{1-r}\right) + o\left(\frac{1}{1-r} \log \frac{1}{1-r}\right),$$

or

$$\sum_1^n (s_n^* - s)^2 r^n = o\left(\frac{1}{1-r} \log \frac{1}{1-r}\right).$$

Putting  $r = 1 - 1/n$  yields

$$\sum_1^n (s_n^* - s)^2 = o(n \log n), \quad \text{as } n \rightarrow \infty,$$

which proves Theorem II.

Addendum (May 27, 1940): To complete the proof of the criterion in §3 we remark that

$$\int_0^t \{\phi(t) \int_0^t \phi(u) \cot \frac{u}{2} R_n(t, u) du\} \cot \frac{t}{2} dt = o(n^3).$$

This follows easily from the fact that strong summability at a point is a local property of the function. A similar remark holds for the criterion of §4. To prove Theorems I and II we could also confine ourselves to the case  $\delta = \pi$ .

I have learned from Mathematical Reviews, vol. 1 (1940), p. 139, that T. Kawata (Proceedings of the Imperial Academy, Tokyo, vol. 15 (1939), pp. 243-246) also gave a simpler proof of Theorem II.

#### LITERATURE

1. L. Fejér, *Zur Summabilitätstheorie der Fourierschen und Laplaceschen Reihe*, Proceedings of the Cambridge Philosophical Society, vol. 34 (1938), pp. 503-509.
2. G. H. Hardy and J. E. Littlewood, *The strong summability of Fourier series*, Fundamenta Mathematicae, vol. 25 (1935), pp. 162-189.
3. Otto Szász, *Selected Topics in Function Theory of a Complex Variable*, Brown University Lectures, 1934-1935.
4. Otto Szász, *On the Cesàro and Riesz means of Fourier series*, Compositio Mathematica, vol. 7 (1939), pp. 112-122.
5. A. Zygmund, *Trigonometrical Series*, 1935.

UNIVERSITY OF CINCINNATI,  
CINCINNATI, OHIO.

## THEORY OF REDUCTION FOR ARITHMETICAL EQUIVALENCE

BY  
HERMANN WEYL

### INTRODUCTION

Minkowski's *Geometrie der Zahlen* as it was published in 1896 led up to two fundamental inequalities concerning a symmetric convex body in relationship to a lattice; in his notation

$$(1) \quad M^n V \leq 2^n$$

and

$$(2) \quad S_1 \cdots S_n V \leq 2^n.$$

The second inequality, which generalizes the first, is a decisive step towards a theory of reduction of arbitrary gauge functions under arithmetical equivalence. In fact the problem of reduction for quadratic forms of  $n$  variables (ellipsoids) was the starting point of Minkowski's investigations. But he must have found that the new instrument which he invented and of which he made so many beautiful applications in other directions was not quite adequate to the goal for which it had originally been devised. For 14 years later he came out with a paper on "Diskontinuitätsbereich für arithmetische Äquivalenz" [1] which makes no use whatsoever of his own geometric methods. This was probably due to two difficulties: he failed to see a way of passing from pseudo-reduction to true reduction for an arbitrary convex body, and in the special case of ellipsoids he found the inequality of true reduction tied up with the selection of a finite number among the linear inequalities which characterize a reduced form. The latter knot was unraveled by a kind of topological argument in a joint paper by L. Bieberbach and I. Schur [2] while K. Mahler in 1938 made an almost trivial remark which removed the first difficulty [3]. In a general overhauling of the geometry of numbers [4], to which the author was led by preparing an introductory talk for a seminar on the subject, he generalized (2) in such a way as to make the approach to that inequality more natural [5], rediscovered Mahler's observation, substituted a simpler argument for that used by Bieberbach and Schur and finally extended Minkowski's second theorem of finiteness. Without this extension certain primitive questions about the topological pattern of equivalent cells would be unanswerable. In a previous paper R. Remak had considerably shortened and sharpened Minkowski's estimate for the coefficients  $\beta_{ij}$  which appear in

---

Presented to the Society, February 24, 1940; received by the editors February 16, 1940.

the Jacobi transformation of a reduced quadratic form [6]. The author found that a considerable part of the theory of reduction could be carried through along the lines of Mahler's approach for arbitrary convex bodies and that this more general procedure results in stronger rather than weaker estimates for the quantities on which the question of finiteness depends.

The present paper sets forth the whole theory *ab ovo*, and hence is partly of a didactic nature; as far as possible it follows the geometric approach dealing with arbitrary convex bodies. In order to prevent it from becoming too dull reading, I have extended the theory to vectors and lattices and forms in which complex numbers or quaternions take the place of real numbers. Chapter I deals with the general theory, Chapter II with the special case of quadratic, Hermitian and "Hamiltonian" forms<sup>(1)</sup>.

## CHAPTER I. GENERAL THEORY OF REDUCTION

### A. THE REAL CASE

1. **Known facts about lattices.** In the  $n$ -dimensional vector space  $E_n$  whose elements are the  $n$ -uples  $\mathbf{x} = (x_1, \dots, x_n)$  of real numbers we consider the lattice  $\mathfrak{L}$  of the vectors with integral components  $x_i$ . The  $n$  unit vectors  $\mathbf{e}_k = (\delta_1^k, \dots, \delta_n^k)$  form a basis of, or span, this lattice in the sense that the lattice vectors appear as sums  $\sum x_i \mathbf{e}_i$  with integral coefficients. Here  $\delta_i^k$  are the Kronecker  $\delta$ 's. Any basis  $\mathfrak{g}_k = (s_1^k, \dots, s_n^k)$  of the lattice arises from the absolute basis  $\mathbf{e}_k$  by a unimodular transformation  $S = \|s_i^k\|$ :

$$\mathfrak{g}_k = \sum_i s_i^k \mathbf{e}_i.$$

The corresponding coordinates,  $x_i$  and  $x'_i$ ,  $\mathbf{x} = \sum x_i \mathbf{e}_i = \sum_k x'_k \mathfrak{g}_k$ , are linked by the equations<sup>(2)</sup>

$$x_i = \sum_k x'_k s_i^k \text{ or briefly, } \mathbf{x} = \mathbf{x}' S.$$

The coefficients  $s_i^k$  are integers and their determinant is  $\pm 1$ . The substitutions  $S$  with these properties form a group  $\{S\}$ , the *modular group*. Our viewpoint is that the vector space is endowed with the lattice, but that the choice of the lattice basis is arbitrary.

<sup>(1)</sup> A brief and masterly treatment of the reduction of quadratic forms along purely arithmetical lines is to be found in a recent paper by C. L. Siegel, *Abhandlungen aus dem mathematischen Seminar der Hansischen Universität*, vol. 13 (1939), pp. 209-239, of which I received a reprint on March 20, 1940. (The number of the journal itself has not yet reached Princeton.) But even against Siegel's highly simplified arithmetical treatment, the geometrical approach retains the advantage of yielding sharper estimates. Siegel has a generalization of the second theorem of finiteness, different from ours, which leads to important applications in the domain of rational indefinite forms. (Added March 25, 1940.)

<sup>(2)</sup> In preparation for a later generalization to quaternions we take good care to put factors in their proper order.



Any  $k$  linearly independent vectors  $b_1, \dots, b_k$  ( $0 \leq k \leq n$ ) span a  $k$ -dimensional subspace

$$E_k = E = [b_1, \dots, b_k].$$

If they are lattice vectors, then  $E$  is a *lattice subspace*  $E_0$  consists of the vector zero only.

A vector  $a$  not in  $E$  may be adjoined to  $E$  and then gives rise to the  $(k+1)$ -dimensional manifold  $E' = [E, a]$  consisting of all sums

$$(3) \quad \xi' = \xi + x a$$

with  $\xi$  in  $E$ ,  $x$  a number. If  $E$  is a lattice subspace and  $a$  a lattice vector, the adjunction is said to be *primitive* provided every lattice vector (3) in  $E'$  has an integral coefficient  $x$  (and hence a lattice component  $\xi$  in  $E$ ).

Suppose  $b_1, \dots, b_k$  are  $k$  linearly independent lattice vectors spanning the lattice subspace  $E = [b_1, \dots, b_k]$ .

LEMMA 1. *There exists a positive integer  $M$  such that every lattice vector in  $E$  is of the form*

$$\frac{y_1}{M} b_1 + \dots + \frac{y_k}{M} b_k$$

where the  $y$ 's are integers.

There are two essentially different proofs of this fact, one resting on divisibility and determinants, the other on considerations of magnitude. The first proof runs as follows. We can select  $n-k$  among the unit vectors  $e_1, \dots, e_n$ , say  $e'_1, \dots, e'_{n-k}$ , such that

$$(4) \quad b_1, \dots, b_k, e'_1, \dots, e'_{n-k}$$

are linearly independent. The determinant of the components of (4) is non-zero; denote its absolute value by  $M$ . Writing down the equation

$$(5) \quad \xi = y_1 b_1 + \dots + y_k b_k + x'_1 e'_1 + \dots + x'_{n-k} e'_{n-k}$$

for any lattice vector  $\xi$  in terms of absolute components, one finds the coefficients  $y$  and  $x'$  to be fractions with the common denominator  $M$ . This applies in particular to the lattice vectors in  $E$  for which  $x'_1 = \dots = x'_{n-k} = 0$ .

The other proof compares  $\mathfrak{L} \cap E = \mathfrak{L}_k$ , "the lattice in  $E$ ," with the coarser lattice  $\mathfrak{L}_k^0$  consisting of all integral combinations of  $b_1, \dots, b_k$ ,

$$(6) \quad y_1 b_1 + \dots + y_k b_k \quad (y_1, \dots, y_k \text{ integers}).$$

We maintain that there is only a finite number  $M$  of vectors in  $\mathfrak{L}_k$  which are incongruent modulo  $\mathfrak{L}_k^0$ . For every vector  $\xi$  in  $E$  there exists a reduced one

$$(7) \quad \xi^* \equiv \xi \pmod{\mathfrak{L}_k^0}, \quad \xi^* = y_1^* b_1 + \dots + y_k^* b_k,$$

which satisfies the inequalities

$$(8) \quad |y_1^*| \leq \frac{1}{2}, \dots, |y_k^*| \leq \frac{1}{2}.$$

Using again the absolute components one readily derives from (8) upper bounds for the  $|x_i^*|$  of any reduced vector  $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ . Hence if the  $x_i^*$  are required to be integers, which is the case when  $\mathbf{x}$  and thus  $\mathbf{x}^*$  is a lattice vector, one finds oneself restricted to a finite number of possibilities. Our result states that the additive Abelian group  $\mathfrak{L}_k/\mathfrak{L}_k^0$  is of finite order  $M$ , and therefore every vector  $\mathbf{x}$  of  $\mathfrak{L}_k$  satisfies the congruence  $M\mathbf{x} \equiv 0 \pmod{\mathfrak{L}_k^0}$ , which was to be proved.

The vectors  $b_1, \dots, b_k$  form a lattice basis of  $E$  if  $\mathfrak{L}_k$  coincides with  $\mathfrak{L}_k^0$ , that is to say, if every lattice vector in  $E$  is of the form (6).

The vector  $\mathfrak{s}_k$  of any basis  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  of  $\mathfrak{L}$  evidently is a primitive adjunction to  $[\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}]$ . More generally, we have

LEMMA 2. Suppose  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  constitute a basis of  $\mathfrak{L}$ . The vector

$$\mathfrak{a} = a_1\mathfrak{s}_1 + \dots + a_n\mathfrak{s}_n$$

is a primitive adjunction to  $E = [\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}]$  if and only if  $a_1, \dots, a_n$  are integers and  $a_k, \dots, a_n$  are without common divisor.

Proof. 1. If  $(a_k, \dots, a_n)$  have a common divisor  $d > 1$ , then

$$(9) \quad \frac{1}{d} (a_k\mathfrak{s}_k + \dots + a_n\mathfrak{s}_n)$$

evidently is a vector  $\mathbf{x}'$  in  $E' = [E, \mathfrak{a}]$  for which the  $x$  in (3) is  $1/d$  and thus not an integer.

2. If one denotes by  $x'_i$  the components of  $\mathbf{x}'$  in (3) with respect to the basis  $\mathfrak{s}_i$ , one has

$$(10) \quad x'_k = xa_k, \dots, x'_n = xa_n.$$

Hence (10) must be integers for any lattice vector  $\mathbf{x}'$  in  $E'$ . However if  $a_k, \dots, a_n$  are without common divisor one can ascertain integers  $l_k, \dots, l_n$  satisfying the equation

$$a_k l_k + \dots + a_n l_n = 1.$$

The integrity of (10) then results in the integrity of

$$x = x'_k l_k + \dots + x'_n l_n$$

itself.

LEMMA 3. Suppose  $E'$  is a given lattice subspace and  $\mathfrak{b}$  a lattice vector outside  $E'$ . Then one can pass from  $E'$  to  $E = [E', \mathfrak{b}]$  by a primitive adjunction  $\mathfrak{s}$ .

Proof. Let  $E$  be spanned by the  $k-1$  linearly independent lattice vectors

$\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}$  and use the notations  $\mathfrak{L}_k, \mathfrak{L}_k^0$  with respect to the basis  $(\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}, \mathfrak{b})$  of  $E$ . We write each vector  $\mathfrak{x}$  of  $\mathfrak{L}_k$  in the form (3),

$$(11) \quad \mathfrak{x} = x\mathfrak{b} + \mathfrak{x}' \quad (\mathfrak{x}' \text{ in } E').$$

If  $M$  is the order of the additive Abelian group  $\mathfrak{L}_k/\mathfrak{L}_k^0$ , we know that

$$(12) \quad Mx = y$$

is an integer. Select a full system of residues

$$\mathfrak{x}^{(0)} = 0, \mathfrak{x}^{(1)}, \dots, \mathfrak{x}^{(M-1)}$$

of  $\mathfrak{L}_k$  modulo  $\mathfrak{L}_k^0$  and denote by  $y^{(0)}=0, y^{(1)}, \dots, y^{(M-1)}$  the corresponding numbers  $y$  as defined by (11), (12). The integers  $M, y^{(1)}, \dots, y^{(M-1)}$  have a greatest common divisor (G.C.D.)  $m^*$ , namely a common divisor expressible as a linear combination

$$lM + l^{(1)}y^{(1)} + \dots + l^{(M-1)}y^{(M-1)}$$

with integral coefficients  $l$ . By forming the corresponding combination

$$\mathfrak{s} = l\mathfrak{b} + l^{(1)}\mathfrak{x}^{(1)} + \dots + l^{(M-1)}\mathfrak{x}^{(M-1)}$$

we obtain a vector  $\mathfrak{s}$  of  $\mathfrak{L}_k$ ,

$$\mathfrak{s} = (m^*/M)\mathfrak{b} + \mathfrak{s}' \quad (\mathfrak{s}' \text{ in } E'),$$

such that for every  $\mathfrak{x}$  in  $\mathfrak{L}_k$  the coefficient  $y$  is divisible by  $m^*$ . This  $\mathfrak{s}$  evidently satisfies our lemma.

Since  $m^*$  is a divisor of  $M$ ,  $M = mm^*$ , we have

$$(13) \quad \mathfrak{s} = (1/m)\mathfrak{b} + t_1\mathfrak{s}_1 + \dots + t_{k-1}\mathfrak{s}_{k-1}.$$

$m$  is a positive integer. Moreover one can assume

$$(14) \quad |t_1| \leq \frac{1}{2}, \dots, |t_{k-1}| \leq \frac{1}{2}.$$

In the special case  $m=1$  one may simply take  $\mathfrak{s} = \mathfrak{b}$ .

We shall use our lemma only for the case when  $\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}$  constitute a lattice basis of  $E'$ . Then the lemma makes possible, by induction with respect to  $k$ , the construction of a lattice basis for any given lattice subspace.

All these simple facts about lattices are well known to the mathematician and the crystallographer. We had to restate them for later use and generalizations.

**2. Gauge functions. Minkowski's inequality.** According to Minkowski, a real-valued continuous function  $f(\mathfrak{x}) = f(x_1, \dots, x_n)$  in vector space is said to be a *gauge function* under the following three conditions:

- (i)  $f(x_1, \dots, x_n) > 0$ , except for  $x_1 = \dots = x_n = 0$ ;
- (ii)  $f(tx_1, \dots, tx_n) = |t| \cdot f(x_1, \dots, x_n)$  for any real factor  $t$ ;
- (iii)  $f(x_1 + x'_1, \dots, x_n + x'_n) \leq f(x_1, \dots, x_n) + f(x'_1, \dots, x'_n)$ .

One may use this function to endow the  $n$ -dimensional affine point space with a metric by ascribing the distance  $f(\overline{pp'})$  to any two points  $p, p'$ . The gauge body  $\mathfrak{R}$  defined by  $f(\mathfrak{x}) < 1$  is an open convex bounded set surrounding the origin  $\mathfrak{x} = 0$ . (Boundedness follows from the fact that  $f(x_1, \dots, x_n)$  has a positive minimum on the sphere  $x_1^2 + \dots + x_n^2 = 1$ .)  $\mathfrak{R}$  has a Jordan volume  $V$ .

Equation (13), together with (14) and  $m \geq 1$ , results in the inequality

$$(15) \quad f(\mathfrak{s}) \leq f(\mathfrak{b}) + \frac{1}{2}\{f(\mathfrak{s}_1) + \dots + f(\mathfrak{s}_{k-1})\}.$$

If one makes the distinction  $m = 1$  or  $m \geq 2$  one finds that  $f(\mathfrak{s})$  cannot exceed both numbers

$$f(\mathfrak{b}), \quad \frac{1}{2}f(\mathfrak{b}) + \frac{1}{2}f(\mathfrak{s}_1) + \dots + \frac{1}{2}f(\mathfrak{s}_{k-1}).$$

Therefore we may state this

SUPPLEMENT TO LEMMA 3. *The vector  $\mathfrak{s}$  may be chosen so that (15) holds, or even so that*

$$(16) \quad f(\mathfrak{s}) \leq \max \{f(\mathfrak{b}), \frac{1}{2}f(\mathfrak{b}) + \frac{1}{2}f(\mathfrak{s}_1) + \dots + \frac{1}{2}f(\mathfrak{s}_{k-1})\}.$$

Minkowski determines a sequence of lattice vectors  $\mathfrak{b}_1, \dots, \mathfrak{b}_n$  and lattice subspaces  $E_0, E_1, \dots, E_n$  starting with the zero-space  $E_0$  by the following induction with respect to  $k$ .

Among all lattice vectors  $\mathfrak{a}$  outside  $E_{k-1}$ , one chooses one,  $\mathfrak{b}_k$ , for which  $f(\mathfrak{a})$  takes on the least possible value, so that  $f(\mathfrak{a}) \geq f(\mathfrak{b}_k)$  for every  $\mathfrak{a}$  outside  $E_{k-1}$ . The space  $E_k$  arises from  $E_{k-1}$  by the adjunction of  $\mathfrak{b}_k$ ,  $E_k = [E_{k-1}, \mathfrak{b}_k]$ .

We put  $f(\mathfrak{b}_k) = M_k$ . Evidently

$$M_1 \leq M_2 \leq \dots \leq M_n.$$

Consider the continuous series of homothetic solids

$$\mathfrak{R}(q): \quad f(\mathfrak{x}) < q$$

increasing with the positive parameter  $q$ . Our  $M_k$  can be described thus:  $\mathfrak{R}(q)$  contains less than  $k$  linearly independent lattice vectors as long as  $q \leq M_k$ , but at least  $k$  such vectors if  $q > M_k$ . Hence  $M_1, \dots, M_n$  are uniquely determined. About these consecutive minima Minkowski proved the fundamental inequality:

THEOREM 1.

$$(2) \quad M_1 \cdots M_n V \leq 2^n.$$

For later purposes we repeat this proposition in the following slightly modified form: Suppose  $M'_1, \dots, M'_n$  are given positive numbers such that the number of linearly independent lattice vectors  $\mathfrak{x}$  for which  $f(\mathfrak{x}) < M'_k$  is less than  $k$ . Then

$$(17) \quad M'_1 \cdots M'_n V \leq 2^n.$$

While  $M_1, \dots, M_n$  are uniquely determined, there may be a certain amount of free play in the choice of  $b_1, \dots, b_n$ . The most one can say about it in general terms is this:

**THEOREM 2.** *If  $b'_1, \dots, b'_n$  are a second set of lattice vectors determined just like  $b_1, \dots, b_n$ , and if, for a certain  $k$ ,  $M_k < M_{k+1}$ , then  $b'_1, \dots, b'_k$  are linear combinations of  $b_1, \dots, b_k$  only.*

**Proof.** Suppose one of the vectors  $b'_1, \dots, b'_k$ , say  $b'_i$ , is not a linear combination of  $b_1, \dots, b_k$ . Then  $b_1, \dots, b_k, b'_i$  are linearly independent, and hence not all the  $k+1$  numbers

$$f(b_1) = M_1, \dots, f(b_k) = M_k, f(b'_i) = M_i$$

can be less than  $M_{k+1}$ . This contradicts the assumption  $M_k < M_{k+1}$ .

The problem of reduction consists in constructing a basis for the lattice  $\mathfrak{L}$  in terms of the given gauge function  $f$ . The vectors  $b_1, \dots, b_n$  do not yet solve the problem because in general they do not span the whole lattice  $\mathfrak{L}$ . Our next task will be to pass from this pseudo-reduction to true reduction, a step well prepared by the considerations of §1.

**3. Reduction.** The only modification needed in the definition of  $b_k$  is the insertion at its proper place of the word "primitive." The new inductive definition of lattice vectors  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  and lattice subspaces  $E_0, E_1, \dots, E_n$  runs as follows:

*Among all primitive adjunctions  $a$  to  $E_{k-1}$ , we choose one,  $b_k$ , for which  $f(a)$  assumes the least possible value, so that*

$$f(a) \geq f(b_k)$$

*for every primitive adjunction  $a$  to  $E_{k-1}$ . Moreover*

$$E_k = [E_{k-1}, b_k].$$

Lemma 3 guarantees the existence of primitive adjunctions  $a$  to  $E_{k-1}$ . We realize by induction that  $\mathfrak{s}_1, \dots, \mathfrak{s}_k$  is a lattice basis for  $E_k$ , hence  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  for the whole space. We put  $f(\mathfrak{s}_k) = L_k$ . Taking Lemma 2 into account, we can give our definition of a reduced basis  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  the following turn:

An  $n$ -uple of integers  $(x_1, \dots, x_n)$  is said to belong to  $X_k$  if  $x_k, \dots, x_n$  are without common divisor. The basis  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  of  $\mathfrak{L}$  is reduced with respect to  $f$ , if for every  $k = 1, \dots, n$  and every  $(x_1, \dots, x_n)$  of  $X_k$  the inequality

$$(18) \quad f(x_1 \mathfrak{s}_1 + \dots + x_n \mathfrak{s}_n) \geq f(\mathfrak{s}_k)$$

holds [7]. Our procedure has led up to this result:

**THEOREM 3.** *For every gauge function  $f$  there exists a reduced basis  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  of the lattice.*

Relation (18) implies

$$f(\mathfrak{s}_{k+1}) \geq f(\mathfrak{s}_k)$$

or

$$(19) \quad L_1 \leq L_2 \leq \cdots \leq L_n.$$

The following proposition ties up pseudo-reduction with the reduction just defined [8]:

**THEOREM 4** (Mahler's theorem). *One has*

$$(20) \quad L_k \leq \theta_k M_k$$

where  $\theta_k$  is a constant independent of the gauge function  $f$ .

An immediate corollary derived from it by Minkowski's inequality (2) is

**THEOREM 5.** *The relation*

$$(21) \quad L_1 \cdots L_n V \leq \mu_n$$

holds with  $\mu_n = 2^n \cdot \theta_1 \theta_2 \cdots \theta_n$ .

**Proof.** After we have ascertained  $\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}$  we determine a primitive adjunction  $\mathfrak{s}$  to  $E' = [\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}]$  by the construction of Lemma 3, choosing  $\mathfrak{b}$  in this particular fashion: One of the  $k$  linearly independent vectors  $\mathfrak{b}_1, \dots, \mathfrak{b}_k$  occurring in Minkowski's construction, say  $\mathfrak{b}_i$ , lies outside  $E'$ . We take  $\mathfrak{b} = \mathfrak{b}_i$  and then find a primitive adjunction  $\mathfrak{s}$  to  $E'$  such that  $[E', \mathfrak{s}] = [E', \mathfrak{b}]$ . By the supplement to Lemma 3 one will have

$$f(\mathfrak{s}) \leq f(\mathfrak{b}) + \frac{1}{2}\{f(\mathfrak{s}_1) + \cdots + f(\mathfrak{s}_{k-1})\}.$$

Since  $f(\mathfrak{b})$  is one of the numbers  $M_1, \dots, M_k$  and hence is less than or equal to  $M_k$ , and since by definition  $L_k \leq f(\mathfrak{s})$ , we find

$$L_k \leq M_k + \frac{1}{2}(L_1 + \cdots + L_{k-1}),$$

which under the assumption of the inequalities

$$L_1 \leq \theta_1 M_1, \dots, L_{k-1} \leq \theta_{k-1} M_{k-1}$$

leads on to

$$L_k \leq \theta_k M_k$$

with

$$(22) \quad \theta_k = 1 + \frac{1}{2}(\theta_1 + \cdots + \theta_{k-1}).$$

Hence Theorem 4 is proved inductively, and, by the recursive relations (22) or

$$\theta_1 = 1; \quad \theta_{k+1} = 1 + \frac{1}{2}(\theta_1 + \cdots + \theta_{k-1} + \theta_k) = \theta_k + \frac{1}{2}\theta_k = \frac{3}{2}\theta_k,$$



we find the following explicit expressions for  $\theta_k$  and  $\mu_n$ :

$$\theta_k = \left(\frac{3}{2}\right)^{k-1}, \quad \mu_n = \left(\frac{3}{2}\right)^{n(n-1)/2}.$$

Suppose  $p_0, p_1, \dots, p_n$  are given numbers satisfying the following conditions:

$$(23) \quad 1 = p_0 \leq p_1 \leq \dots \leq p_n.$$

A basis  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_n$  of  $\mathfrak{L}$  is said to have the property  $B(p_1, \dots, p_n)$  if the inequality

$$f(x_1 \mathfrak{s}'_1 + \dots + x_n \mathfrak{s}'_n) \geq (1/p_k) f(\mathfrak{s}_k)$$

holds whenever  $(x_1, \dots, x_n)$  is an  $n$ -uple in  $X_k$  and  $k$  one of the indices  $1, \dots, n$ . By exploiting our method to the full we arrive at the following [9] generalization of Theorem 4:

**THEOREM 6.** *If the lattice basis  $\mathfrak{s}'_k$  has the property  $B(p_1, \dots, p_n)$ , then the values  $f(\mathfrak{s}'_k) = L'_k$  satisfy the inequalities*

$$(24) \quad L'_k \geq \frac{1}{p_i} L'_i \quad (\text{for } k > i)$$

and

$$(25) \quad L'_k \leq \theta_k(p) \cdot M_k \quad (k = 1, \dots, n)$$

with a constant  $\theta_k(p)$  depending on  $p_1, \dots, p_k$  but not on  $f$ .

Relation (24) is a consequence of the fact that  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$  is an  $n$ -uple in  $X_i$  if  $k > i$ . Otherwise the proof follows the same road as before. (22) gives place to this recursive equation:

$$\theta_k(p)/p_k = 1 + \frac{1}{2}(\theta_1(p) + \dots + \theta_{k-1}(p))$$

which in the same manner readily leads to

$$\theta_k(p) = p_k \cdot \prod_{i=1}^{k-1} (1 + \frac{1}{2} p_i).$$

One sees that  $\theta_k(p)/p_k$  increases with  $k$ , and therefore (23) implies

$$(26) \quad 1 = \theta_0(p) \leq \theta_1(p) \leq \dots \leq \theta_n(p).$$

One can repeat our whole argument after replacing (15) by the sharper and slightly more complex inequality (16). One then obtains this

**SUPPLEMENT TO THEOREMS 4-6.** *One may choose*

$$(27) \quad \theta_k = \left(\frac{3}{2}\right)^{k-1}, \quad \mu_n = \left(\frac{3}{2}\right)^{n(n-1)/2}, \quad \theta_k(p) = p_k \cdot \prod_{i=1}^{k-1} (1 + \frac{1}{2} p_i),$$

or, with a slight improvement,

$$(28) \quad \begin{aligned} \theta_1 &= 1, & \theta_k &= \left(\frac{3}{2}\right)^{k-2} \quad (\text{for } k \geq 2); & \mu_n &= \left(\frac{3}{2}\right)^{(n-1)(n-2)/2}; \\ \theta_1(p) &= p_1, & \theta_k(p) &= p_k \cdot \frac{1+p_1}{2} \cdot \prod_{i=2}^{k-1} (1 + \frac{1}{2}p_i) \quad (\text{for } k \geq 2). \end{aligned}$$

Shifting the accent, we call a gauge function  $f(x_1, \dots, x_n)$  reduced if it satisfies the inequalities

$$f(x_1, \dots, x_n) \geq f(\delta_1^k, \dots, \delta_n^k)$$

for any vector  $(x_1, \dots, x_n)$  in  $X_k$  and  $k=1, \dots, n$ . This means that the unit vectors  $e_k = (\delta_1^k, \dots, \delta_n^k)$  form a reduced lattice basis with respect to  $f$ . The inequalities (20) then hold for  $L_k = f(e_k)$ . If  $f(x)$  is any gauge function and  $\theta_1, \dots, \theta_n$  a reduced lattice basis with respect to  $f$ , we may set

$$f(x_1\theta_1 + \dots + x_n\theta_n) = f^*(x_1, \dots, x_n).$$

Then  $f^*(x_1, \dots, x_n)$  is a reduced gauge function, and we see that any gauge function  $f$  can be carried over into a reduced one by a unimodular transformation  $S$  of its variables. We shall adopt this terminology in Chapter II while at present we stick to talking in terms of reduced bases rather than gauge functions.

4. **The question of uniqueness.** Denote by  $X_k^*$  the set  $X_k$  after excluding the two  $n$ -uples

$$(x_1, \dots, x_n) = \pm (\delta_1^k, \dots, \delta_n^k).$$

The lattice basis  $\theta_1, \dots, \theta_n$  is said to be *properly reduced* when for every  $k=1, \dots, n$  and for every  $(x_1, \dots, x_n)$  in  $X_k^*$  the inequality (18) holds with the  $>$  sign.

The  $2^n$  diagonal transformations of the modular group,

$$J: \theta_1' = \pm \theta_1, \dots, \theta_n' = \pm \theta_n$$

(all possible combinations of signs admitted) form a finite Abelian subgroup  $\{J\}$  of order  $2^n$ . Its generators are the involutions  $J_1, \dots, J_n$  which change one sign at a time:

$$J_k: \theta_k' = -\theta_k \text{ and } \theta_j' = \theta_j \text{ for all } j \neq k.$$

Clearly the  $J$  carry a reduced basis  $(\theta_1, \dots, \theta_n)$  into a reduced one. The first result concerning the question of uniqueness is that this exhausts the possibilities, provided  $(\theta_1, \dots, \theta_n)$  is *properly reduced* [10]. Of two lattice bases  $(\theta_1, \dots, \theta_n)$  and  $(\theta_1', \dots, \theta_n')$ , the first is called *lower* than the second provided the first nonvanishing difference

$$f(\theta_1') - f(\theta_1), \dots, f(\theta_n') - f(\theta_n)$$

happens to be positive (which includes the case for which they are all zero).

**THEOREM 7.** Let  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$  be any lattice basis and  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  be a properly reduced lattice basis. In these circumstances  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  is lower than  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$ , and the equations

$$f(\mathfrak{s}'_1) = f(\mathfrak{s}_1), \dots, f(\mathfrak{s}'_k) = f(\mathfrak{s}_k)$$

imply

$$\mathfrak{s}'_1 = \pm \mathfrak{s}_1, \dots, \mathfrak{s}'_k = \pm \mathfrak{s}_k.$$

If  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$  is reduced and  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  is properly reduced, then

$$\mathfrak{s}'_1 = \pm \mathfrak{s}_1, \dots, \mathfrak{s}'_n = \pm \mathfrak{s}_n.$$

**Proof.** Under the hypothesis that  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  is properly reduced, we have to show that

$$(29) \quad \mathfrak{s}'_1 = \pm \mathfrak{s}_1, \dots, \mathfrak{s}'_{k-1} = \pm \mathfrak{s}_{k-1}$$

imply  $f(\mathfrak{s}'_k) \geq f(\mathfrak{s}_k)$ , and even  $f(\mathfrak{s}'_k) > f(\mathfrak{s}_k)$  unless  $\mathfrak{s}'_k = \pm \mathfrak{s}_k$ .

Because of (29),  $\mathfrak{s}'_k$  is a primitive adjunction to

$$[\mathfrak{s}'_1, \dots, \mathfrak{s}'_{k-1}] = [\mathfrak{s}_1, \dots, \mathfrak{s}_{k-1}],$$

and hence

$$(30) \quad f(\mathfrak{s}'_k) \geq f(\mathfrak{s}_k).$$

As  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  is properly reduced, the equality sign in (30) will hold only if  $\mathfrak{s}'_k = \pm \mathfrak{s}_k$ .

Suppose  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_n$  is reduced and (29) holds. Since  $\mathfrak{s}_k$  is a primitive adjunction to  $[\mathfrak{s}'_1, \dots, \mathfrak{s}'_{k-1}]$ , we must have  $f(\mathfrak{s}_k) \geq f(\mathfrak{s}'_k)$  in addition to (30), and hence  $f(\mathfrak{s}'_k) = f(\mathfrak{s}_k)$ , an equation which we have just found impossible unless  $\mathfrak{s}'_k = \pm \mathfrak{s}_k$ . This establishes the full content of our theorem.

Much less can be said if the reduced basis  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  is not properly reduced.

**THEOREM 8.** If

$$\mathfrak{s}_1, \dots, \mathfrak{s}_n; \quad \mathfrak{s}'_1, \dots, \mathfrak{s}'_n$$

are two reduced bases, then

$$L_k = f(\mathfrak{s}_k), \quad L'_k = f(\mathfrak{s}'_k)$$

satisfy the inequalities

$$(31) \quad \theta_k L_k \geq L'_k, \quad \theta_k L'_k \geq L_k.$$

(This proposition indicates how far the uniqueness of the  $M_k$  survives for the  $L_k$ .)

**Proof.** Because there are  $k$  linearly independent lattice vectors  $\mathfrak{r} = \mathfrak{s}_1, \dots, \mathfrak{s}_k$  for which  $f(\mathfrak{r}) \leq L_k$ ,  $L_k$  cannot be smaller than  $M_k$ . Hence

$$(32) \quad \begin{aligned} M_k &\leq L_k, & L_k &\leq \theta_k M_k; \\ M_k &\leq L'_k, & L'_k &\leq \theta_k M_k. \end{aligned}$$

Elimination of  $M_k$  leads to the two inequalities (31).

The case when  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  is reduced while the basis  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$  has the property  $B(p_1, \dots, p_n)$  will also be needed later. The  $k$  linearly independent vectors  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_{k-1}, \mathfrak{s}'_k$  impart values to  $f$  which are less than or equal to

$$p_1 L'_k, \dots, p_{k-1} L'_k, L'_k$$

respectively. Hence

$$M_k \leq p_{k-1} L'_k, \quad L'_k \leq \theta_k(p) \cdot M_k.$$

Substituting these inequalities for the second line of (32) and again eliminating  $M_k$  we find:

**THEOREM 8<sub>p</sub>.** For a reduced basis  $(\mathfrak{s}_1, \dots, \mathfrak{s}_n)$  and a basis  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$  of the property

$$B(p_1, \dots, p_n) \quad (1 = p_0 \leq p_1 \leq \dots \leq p_n)$$

the values

$$L_k = f(\mathfrak{s}_k), \quad L'_k = f(\mathfrak{s}'_k)$$

satisfy the inequalities

$$(33) \quad L'_k \leq \theta_k(p) \cdot L_k, \quad L_k \leq \theta_k p_{k-1} \cdot L'_k.$$

With the same effort one could have established similar relations for two bases of the properties  $B(p_1, \dots, p_n)$  and  $B(p'_1, \dots, p'_n)$  respectively. The present generality, however, is sufficient for our purposes.

**THEOREM 9<sub>p</sub>.** If, for a certain  $k = 1, \dots, n-1$ ,

$$(34) \quad \theta_k(p) \theta_{k+1} \cdot L_k < L_{k+1},$$

then  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_k$  are linear combinations of the vectors  $\mathfrak{s}_1, \dots, \mathfrak{s}_k$  only and thus arise from them by a unimodular transformation of degree  $k$ .

**Proof.** Suppose that in one of the vectors  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_k$ , say

$$\mathfrak{s}'_i = s_1^i \mathfrak{s}_1 + \dots + s_n^i \mathfrak{s}_n,$$

not all the components  $s_j^i$ , ( $j = k+1, \dots, n$ ), vanish. Then  $\mathfrak{s}_1, \dots, \mathfrak{s}_k, \mathfrak{s}'_i$  are linearly independent and hence the maximum of the  $k+1$  numbers

$$L_1 = f(\mathfrak{s}_1), \dots, L_k = f(\mathfrak{s}_k), L'_i = f(\mathfrak{s}'_i)$$

must be greater than or equal to  $M_{k+1}$ . If on the contrary

$$(35) \quad L_1, \dots, L_k; \quad L'_1, \dots, L'_k$$

are all less than  $M_{k+1}$ , then the  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_k$  are linear combinations of  $\mathfrak{s}_1, \dots, \mathfrak{s}_k$  only. Now

$$L'_i \leq \theta_i(p) \cdot L_i \quad (i = 1, \dots, k),$$

and owing to

$$L_1 \leq \dots \leq L_k, \quad 1 \leq \theta_1(p) \leq \dots \leq \theta_k(p)$$

all our requirements concerning (35) can be met by the one condition

$$\theta_k(p) \cdot L_k < M_{k+1}$$

which in its turn is a consequence of

$$\theta_k(p) \cdot L_k < L_{k+1}/\theta_{k+1}$$

because  $L_{k+1} \leq \theta_{k+1} M_{k+1}$ .

In the particular case where  $(\mathfrak{s}'_1, \dots, \mathfrak{s}'_n)$  is likewise reduced ( $p_1 = \dots = p_n = 1$ ), we have the following close parallel to Theorem 2:

**THEOREM 9.** *Let  $\mathfrak{s}_1, \dots, \mathfrak{s}_n$  and  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_n$  be two reduced bases of  $\mathfrak{L}$ , and  $f(\mathfrak{s}_k) = L_k$ . Suppose that moreover, for some  $k \leq n-1$ ,*

$$\theta_k \theta_{k+1} L_k < L_{k+1}.$$

*Then the first  $k$  vectors  $\mathfrak{s}'_1, \dots, \mathfrak{s}'_k$  are linear combinations of  $\mathfrak{s}_1, \dots, \mathfrak{s}_k$  only.*

#### B. THE IMAGINARY AND QUATERNION CASES

**5. Integers and Minkowski's inequality in the complex field.** *Complex numbers  $\xi = x_0 + ix_1$  have two real components  $x_0, x_1$ . We denote the conjugate by  $\bar{\xi} = x_0 - ix_1$ . Trace and norm:*

$$\text{tr } \xi = \xi + \bar{\xi} = 2x_0, \quad N\xi = \xi\bar{\xi} = |\xi|^2 = x_0^2 + x_1^2$$

are real and the coefficients of a quadratic equation satisfied by  $\xi$ :

$$(36) \quad \xi^2 - \xi \cdot \text{tr } \xi + N\xi = 0.$$

Let  $\omega$  be a non-real number.  $1, \omega$  span a lattice  $\mathcal{Y}$  in the Gaussian plane consisting of all numbers

$$(37) \quad \xi = y_0 + y_1\omega \quad (y_0, y_1 \text{ integers}).$$

If  $\mathcal{Y}$  is closed with respect to multiplication and the operation  $\xi \rightarrow \bar{\xi}$ , then  $\mathcal{Y}$  is a self-conjugate ring, and we agree to call the elements of  $\mathcal{Y}$  *integers*. Owing to the choice of  $1$  as an element of the lattice basis  $1, \omega$ , the only real integers (with  $y_1 = 0$ ) are the common rational integers. Trace and norm of an integer  $\xi$  are rational integers. Hence the quadratic equation (36) for  $\xi = \omega$  shows that  $\omega$

is of the form  $\frac{1}{2}(c+id^{1/2})$  where  $c$  and  $d$  are rational integers and either

$$c \equiv 0 \pmod{2}, \quad d \equiv 0 \pmod{4}, \quad \text{or} \quad c \equiv 1 \pmod{2}, \quad d \equiv 1 \pmod{4}.$$

The lattice  $\mathcal{Y}$  is rectangular in the first, rhombic in the second case. The density of the lattice  $\mathcal{Y}$ , that is to say, the area of its fundamental parallelogram spanned by  $1, \omega$ , is  $\frac{1}{2}d^{1/2}$ .

The numbers of the form (37) with rational coefficients  $y_0, y_1$  form the embedding field  $\mathcal{Y}_0$ . Indeed if  $\xi \neq 0$  is in  $\mathcal{Y}_0$  so is

$$\xi^{-1} = \bar{\xi}/N\xi.$$

$\mathcal{Y}_0$  is the quadratic field over the rational field determined by  $(-d)^{1/2}$ . The  $x_0, x_1$  and  $y_0, y_1$ , formula (37), are always spoken of as the  $x$ - and  $y$ -components of a complex number  $\xi = x_0 + ix_1$ .

We ask for the least radius  $r$  such that the circles of radius  $r$  around all integers cover the whole  $\xi$ -plane. One readily finds in the rectangular case,

$$r = \frac{1}{2}(1 + \frac{1}{4}d)^{1/2},$$

and in the rhombic case

$$r = \frac{1+d}{4d^{1/2}}.$$

If  $\xi$  is any complex number, one can always ascertain an integer  $\alpha$  such that

$$N(\xi - \alpha) \leq r^2.$$

Another constant which will crop up later is the least norm  $e^2$  of an integer  $\alpha \neq 0$  which is not a unit (i.e. for which  $1/\alpha$  is no integer);  $e$  is either  $2^{1/2}, 3^{1/2}$  or 2.

We operate in a vector space  $E_n$  of  $2n$  real dimensions whose vectors  $\xi = (\xi_1, \dots, \xi_n)$  have arbitrary complex coordinates  $\xi_i$ . The lattice  $\mathcal{V}$  consists of all vectors whose coordinates  $\xi_i$  are integers (elements of  $\mathcal{Y}$ ). The notion of a lattice basis needs no explanation. The modular group  $\{S\}$  consists of all unimodular transformations  $S$ ,

$$\xi_i = \sum_k \xi'_k \sigma_i^k$$

with integral coefficients  $\sigma_i^k$  whose determinant is a unit  $\epsilon$ .

A *gauge function* is a real-valued continuous function  $f(\xi_1, \dots, \xi_n)$  with the following three properties:

- (i)  $f(\xi_1, \dots, \xi_n) > 0$  except for  $(\xi_1, \dots, \xi_n) = (0, \dots, 0)$ ;
- (ii)  $f(\tau\xi_1, \dots, \tau\xi_n) = |\tau| \cdot f(\xi_1, \dots, \xi_n)$ ;
- (iii)  $f(\xi_1 + \xi'_1, \dots, \xi_n + \xi'_n) \leq f(\xi_1, \dots, \xi_n) + f(\xi'_1, \dots, \xi'_n)$ .

We introduce real coordinates  $x_{k0}, x_{k1}$  by  $\xi_k = x_{k0} + ix_{k1}$  and use them in defin-



ing the volumes of solids in our space. In particular  $V$  denotes the volume of the gauge body

$$\mathfrak{R}: f(\xi_1, \dots, \xi_n) < 1.$$

We carry out Minkowski's construction according to the same recipe as in the real case and thus determine  $n$  lattice vectors  $b_1, \dots, b_n$  and consecutive minima  $M_k = f(b_k)$ . Our first concern is the analogue of Minkowski's inequality:

THEOREM 1\*.

$$(38) \quad M_1^2 \cdots M_n^2 V \leq (2d^{1/2})^n.$$

We resort to Minkowski's original inequality in the form (17). But under the present circumstances we deal with  $2n$  real coordinates  $x_{k0}, x_{k1}$  and with a lattice which is the direct product of  $n$  two-dimensional lattices of density  $\frac{1}{2}d^{1/2}$  rather than 1. Hence the right side in (17) is to be replaced by

$$2^{2n}(\frac{1}{2}d^{1/2})^n = (2d^{1/2})^n.$$

The only lattice vectors  $\mathfrak{x}$  for which  $f(\mathfrak{x}) < M_k$  are linear combinations of  $b_1, \dots, b_{k-1}$  with complex coefficients. Hence there are at most  $2(k-1)$  vectors satisfying this inequality which are linearly independent in the real sense. Consequently we may take

$$M'_{2k-1} = M'_{2k} = M_k,$$

and in this way the inequality

$$M'_1 \cdots M'_{2n} V \leq (2d^{1/2})^n$$

results in (38).

**6. The same for quaternions.** A quaternion  $\xi$  has four real components  $(x_0, x_1, x_2, x_3)$ . The conjugate is  $\bar{\xi} = (x_0, -x_1, -x_2, -x_3)$ . The quaternions  $(x, 0, 0, 0)$  can be identified with the real numbers  $x$ . Both trace and norm:

$$\text{tr } \xi = \xi + \bar{\xi} = 2x_0, \quad N\xi = \xi\bar{\xi} = |\xi|^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2,$$

are such real numbers. Every quaternion  $\xi \neq 0$  has its reciprocal

$$(39) \quad \xi^{-1} = \bar{\xi}/N\xi;$$

but since multiplication is noncommutative we have to do with a division algebra rather than a field. Each quaternion  $\xi$  satisfies the quadratic equation (36) with real coefficients.

Any lattice  $\mathfrak{Y}$  in the four-dimensional space with the real coordinates  $x_0, x_1, x_2, x_3$  which is spanned by four linearly independent quaternions including 1,

$$(40) \quad \omega_0 = 1, \quad \omega_1, \quad \omega_2, \quad \omega_3,$$

may serve to define the integral quaternions as those of the form

$$(41) \quad \xi = y_0\omega_0 + y_1\omega_1 + y_2\omega_2 + y_3\omega_3$$

with ordinary integral coefficients  $y$ , provided  $\mathcal{Y}$  is closed with respect to multiplication and the operation  $\xi \rightarrow \bar{\xi}$ . Then trace and norm of a quaternion integer are rational integers. As (39) shows, the quaternions (41) with rational  $y$  form the embedding field  $\mathcal{Y}_0$ . We denote by  $\frac{1}{4}d$  the density of the lattice  $\mathcal{Y}$ , and maintain that  $d$  is a rational integer. Although this fact is of little importance to us I shall briefly indicate its proof.

With (41) we form

$$(42) \quad N\xi = \sum_{i,k=0}^3 a_{ik}y_iy_k (= x_0^2 + x_1^2 + x_2^2 + x_3^2).$$

The coefficients

$$(43) \quad a_{ii} \text{ and } 2a_{ik} \text{ for } i \neq k$$

are rational integers. According to the transformation theory of quadratic forms the discriminant of (42) is  $(\frac{1}{4}d)^2$  and hence, because of (43),  $d^2$  is a rational integer. On the other side let us study the field  $\mathcal{Y}_0$  and any basis  $\omega_0 = 1, \omega_1, \omega_2, \omega_3$  of the field. Starting with (40) we may first subtract from  $\omega_1$  and  $\omega_2$  half their traces and thus provide for the conditions

$$\bar{\omega}_1 = -\omega_1, \quad \bar{\omega}_2 = -\omega_2.$$

Then  $\omega_1\omega_2 + \omega_2\omega_1$  is the trace of  $\omega_1\omega_2$  and hence a real rational number  $2c$ . Replacing  $\omega_2$  by  $\omega_2 + c\omega_1$ , one gets

$$\omega_2\omega_1 = -\omega_1\omega_2.$$

$\omega_1\omega_2$  is in the field. Choosing it as  $\omega_3$  the form (42) becomes

$$y_0^2 + ay_1^2 + by_2^2 + aby_3^2$$

which shows that its discriminant is the square of a rational number. This property persists for any basis of  $\mathcal{Y}_0$ . Hence  $d^2$  is the square of a rational number  $d$ , and, as  $d^2$  is integral, so is  $d$  itself [11].

$r$  and  $e$  have the same significance as before.

The vectors  $\xi = (\xi_1, \dots, \xi_n)$  which we now consider have arbitrary quaternions  $\xi_k$  for their components,

$$\begin{aligned} \xi_k &= (x_{k0}, x_{k1}, x_{k2}, x_{k3}) \\ &= y_{k0}\omega_0 + y_{k1}\omega_1 + y_{k2}\omega_2 + y_{k3}\omega_3. \end{aligned}$$

The definition of lattice vectors remains unchanged. The modular group con-

sists of all pairs of mutually inverse transformations

$$\xi_i = \sum_k \xi'_k \sigma_i^k, \quad \xi'_i = \sum_k \xi_k \tau_i^k$$

with integral coefficients  $\sigma_i^k, \tau_i^k$ . (This modification of the definition is forced upon us because a quaternion matrix  $\|\sigma_i^k\|$  has no determinant.) One has to observe carefully the position of the factors. Our convention is that the subspace spanned by  $k$  linearly independent vectors  $b_1, \dots, b_k$  consists of the vectors  $\eta_1 b_1 + \dots + \eta_k b_k$  with the coefficients  $\eta$  in front of the vectors.

The description of a gauge function by the three properties (i), (ii), (iii) stays unaltered, with the factor  $\tau$  in front of the variables  $\xi_1, \dots, \xi_n$  in (ii). Minkowski's inequality assumes the form

$$M_1^4 \cdots M_n^4 \cdot V \leq 2^{4n} \left(\frac{1}{4}d\right)^n,$$

which we put down as

THEOREM 1\*\*.

$$(44) \quad M_1^2 \cdots M_n^2 \cdot V^{1/2} \leq (2d^{1/2})^n.$$

7. **Reduction.** What remains will be done simultaneously for the imaginary and the quaternion cases in such language as applies literally to the more complex of the two. We have to check Lemmas 1-3 of §1 as to their validity under the new circumstances.

Both proofs of Lemma 1 go through with the following precautions. (5) is to be written down in terms of the  $4n$  integral  $y$  components of the vectors and coefficients concerned, and the positive rational integer  $M$  is the absolute value of the determinant of the linear equations with  $4n$  unknowns thus obtained. The inequalities (6) for a reduced vector (7),

$$\xi^* = \eta_1^* b_1 + \dots + \eta_k^* b_k,$$

must be replaced by

$$N\eta_1^* \leq r^2, \dots, N\eta_k^* \leq r^2.$$

In order to secure the validity of Lemmas 2 and 3 an essentially new assumption has to be made:

HYPOTHESIS P. *Every left or right ideal in the ring  $\mathfrak{f}$  is a principal ideal.*

As far as left ideals are concerned it requires: Any integers

$$(\alpha_1, \dots, \alpha_h) \neq (0, \dots, 0)$$

have a left common divisor  $\delta$ ,

$$\alpha_1 = \delta \cdot \beta_1, \dots, \alpha_h = \delta \cdot \beta_h \quad (\beta_1, \dots, \beta_h \text{ integers}),$$

which can be written as a linear combination

$$(45) \quad \alpha_1 \lambda_1 + \cdots + \alpha_h \lambda_h$$

with integral coefficients  $\lambda_i$ . This divisor  $\delta$ , which up to a right unit factor is uniquely determined, is called the left G.C.D. of  $\alpha_1, \dots, \alpha_h$ . (The integers represented by (45) if the  $\lambda_i$  range independently over all integers coincide with the values of  $\delta \cdot \mu$  for all possible integral values of  $\mu$ . It is sufficient to make the requirement for two integers  $\alpha_1, \alpha_2$ .)

Lemma 2, in which the last words "*without common divisor*" must be changed into "*without left common divisor*," is true under the hypothesis P for *left* ideals ( $P_l$ ). Change the Roman into Greek letters and define  $d$ , or rather  $\delta$ , as the left G.C.D. of  $\alpha_1, \dots, \alpha_n$ . The alternative 1 occurs if  $\delta$  is not a unit, the alternative 2 if  $\alpha_1, \dots, \alpha_n$  are without left common divisor (which means, of course, that they have no left common divisors except *units*).

One has merely to glance through the proof of Lemma 3 in order to realize that it depends on the hypothesis P for *right* ideals. We obtain the primitive adjunction in the form

$$\delta = (1/\mu)b + \tau_1 \delta_1 + \cdots + \tau_{k-1} \delta_{k-1}$$

where  $\mu$  is a nonzero integer and the  $\tau$  satisfy the inequalities

$$N\tau_1 \leq r^2, \dots, N\tau_{k-1} \leq r^2.$$

If  $\mu$  is a unit one may take

$$\delta = b, \text{ i.e., } \mu = 1, \quad \tau_1 = \cdots = \tau_{k-1} = 0.$$

As  $N\mu \geq 1$  for any integer  $\mu \neq 0$ , the inequality (15) is turned over into

$$f(\delta) \leq f(b) + r\{f(\delta_1) + \cdots + f(\delta_{k-1})\}$$

while in (16) the smallest norm  $e^2 > 1$  of integers makes its appearance:

$$f(\delta) \leq \max \{f(b), (1/e)f(b) + r(f(\delta_1) + \cdots + f(\delta_{k-1}))\}.$$

Incidentally hypotheses  $P_l$  and  $P_r$  are fulfilled if  $r < 1$ . For then Euclid's algorithm for the G.C.D. goes through. In the complex field this happens for the rectangular lattices  $\mathcal{Y}$  with  $d=4$  (Gaussian field) and  $d=8$ , and for the rhombic lattices  $\mathcal{Y}$  with  $d=3, 7, 11$ . The most important example for quaternions is the classical case first treated by A. Hurwitz [12]: he declares a quaternion  $(x_0, x_1, x_2, x_3)$  to be integral when  $2x_0, 2x_1, 2x_2, 2x_3$  are rational integers either congruent to  $(0, 0, 0, 0)$  or to  $(1, 1, 1, 1)$  modulo 2. One realizes at once that here  $r < 1$ ; the exact value is  $r = 1/3^{1/2}$ .

The whole theory of reduction of §§3 and 4 will now go through, practically without alterations. We indicate the few changes to be made.  $X_k$  is the set of all  $n$ -uples  $(\xi_1, \dots, \xi_n)$  for which  $\xi_1, \dots, \xi_n$  are integral and  $\xi_k, \dots, \xi_n$

without *left* common divisor.  $X_k^*$  arises from  $X_k$  by excluding the following  $n$ -uples:

$$\epsilon(\delta_1^k, \dots, \delta_n^k) \quad (\epsilon \text{ a unit}).$$

$\{J\}$  consists of the diagonal transformations

$$J: \vartheta_1' = \epsilon_1 \vartheta_1, \dots, \vartheta_n' = \epsilon_n \vartheta_n, \text{ or } \xi_1 = \epsilon_1' \epsilon_1, \dots, \xi_n = \epsilon_n' \epsilon_n$$

where  $\epsilon_1, \dots, \epsilon_n$  are units. This group is the direct product of  $n$  factors each of which is isomorphic with the group of units. The most essential point concerns the values of the constants  $\theta_k$ ,  $\theta_k(p)$  and  $\mu_n$ .

Instead of the recursive formula (22) we get  $\theta_k = 1 + r(\theta_1 + \dots + \theta_{k-1})$ , leading to

$$\theta_k = (1 + r)^{k-1}.$$

Similarly

$$\theta_k(p) = p_k \cdot \prod_{i=1}^{k-1} (1 + r p_i).$$

THEOREM 5\*. *The inequality*

$$L_1^2 \cdots L_n^2 \cdot V^{-1/\kappa} \leq \mu_n^2$$

holds, where  $\kappa = 1, 2, 4$  characterize the real, imaginary and quaternion cases respectively and

$$\mu_n^2 = \{2d^{1/2} \cdot (1 + r)^{n-1}\}^n.$$

(In the real case  $d = 4$ ,  $r = \frac{1}{2}$ .)

The same trick as used before, compare formulas (27) and (28), allows us to improve to some extent these values of  $\theta_k$ ,  $\theta_k(p)$  and  $\mu_n$ .

## CHAPTER II. REDUCTION OF QUADRATIC, HERMITIAN AND HAMILTONIAN FORMS

### 8. Jacobi transformation. A quadratic form

$$(46) \quad f(\mathbf{x}) = \sum g_{ij} x_i x_j \quad (i, j = 1, \dots, n)$$

of  $n$  variables  $(x_1, \dots, x_n) = \mathbf{x}$  is characterized by its real symmetric coefficients  $g_{ij} = g_{ji}$  and may thus be denoted by  $f = \{g_{ij}\}$ . All quadratic forms constitute a linear space  $R$  of  $N = \frac{1}{2}n(n+1)$  dimensions. In the imaginary and the quaternion cases the analogues are the *Hermitian* and "*Hamiltonian*" forms respectively,

$$(47) \quad f(\xi_1, \dots, \xi_n) = \sum_{i,j} \xi_i \gamma_{ij} \bar{\xi}_j$$

whose complex or quaternion coefficients satisfy the symmetry condition

$$(48) \quad \gamma_{ji} = \bar{\gamma}_{ij}.$$

The conjugate of a product is the product of the conjugates in inverted order. This rule at once shows that the value of  $f$  is real,  $\bar{f}=f$ . In the quaternion case one has to watch out for the order of the factors on the right side of (47). The substitution  $x_i \rightarrow tx_i$  multiplies the quadratic form (46) with  $t^2 = |t|^2 = Nt$ , while  $\xi_i \rightarrow \xi_i$  changes (47) into  $\tau f \bar{\tau}$ , or since  $f$  is real, into

$$\bar{\tau} \tau \cdot f = N \tau \cdot f.$$

The diagonal coefficients  $\gamma_{ii}$  are real while the skew coefficients  $\gamma_{ij}$  on one side of the diagonal,  $i < j$ , may be chosen arbitrarily and then determine the coefficients  $\gamma_{ji}$  on the other side by (48). Hence the quadratic, Hermitian and Hamiltonian forms  $f$  constitute linear spaces of

$$N = n + \kappa \cdot \frac{n(n-1)}{2}$$

or of

$$N = \frac{1}{2}n(n+1), \quad n^2, \quad n(2n-1)$$

dimensions respectively. The form  $f$  is said to be *positive* if  $f(\mathbf{r}) > 0$  except for  $\mathbf{r} = 0$ . According to our remarks above,  $f^{1,2}$  may then serve as gauge function in the real, imaginary or quaternion vector spaces.

*Jacobi's transformation* is a uniquely determined linear transformation of recursive character of a positive quadratic form into a square sum. It is nothing else than the method of "completing the square" which, probably some 4000 years ago, was invented for the solution of quadratic equations. It no less applies to Hermitian and Hamiltonian forms, though in the latter case we have to bear in mind that there are no determinants. Thus we had better disregard this formal tool altogether. The discriminant of the form will be defined by recursion in the course of our construction. Its general explicit expression in terms of the coefficients  $\gamma_{ij}$  is a task about which we need not bother here [13]. I now give the description of the process for positive Hamiltonian forms  $f$ .

If  $f$  is positive, then  $\gamma_{11}$  is real and greater than 0,  $\gamma_{11} = q_1$ . We form

$$\xi_1 = \xi_1 + \xi_2 \frac{\gamma_{21}}{\gamma_{11}} + \cdots + \xi_n \frac{\gamma_{n1}}{\gamma_{11}}$$

which implies

$$\bar{\xi}_1 = \bar{\xi}_1 + \frac{\gamma_{12}}{\gamma_{11}} \bar{\xi}_2 + \cdots + \frac{\gamma_{1n}}{\gamma_{11}} \bar{\xi}_n$$

and find



$$(49) \quad f(\xi_1, \dots, \xi_n) = q_1 \xi_1^2 + f^*(\xi_2, \dots, \xi_n)$$

where the remainder  $f^*$  depends on the variables  $\xi_2, \dots, \xi_n$  only. Incidentally its coefficients are given by

$$(50) \quad \gamma_{ij}^* = \gamma_{ij} - \frac{\gamma_{i1}\gamma_{1j}}{\gamma_{11}}$$

$f^*$  is positive; for if  $\xi_2, \dots, \xi_n$  are any given values we may determine  $\xi_1$  by the equation

$$\xi_1 + \xi_2 \frac{\gamma_{21}}{\gamma_{11}} + \dots + \xi_n \frac{\gamma_{n1}}{\gamma_{11}} = 0$$

and then

$$f^*(\xi_2, \dots, \xi_n) = f(\xi_1, \xi_2, \dots, \xi_n) > 0$$

except for  $\xi_2 = \dots = \xi_n = 0$ . Iteration of the splitting (49), therefore, leads to an expression

$$(51) \quad f(\mathfrak{x}) = q_1 |\zeta_1|^2 + \dots + q_n |\zeta_n|^2$$

(Jacobi's transform) where the  $q$  are positive numbers and  $\zeta_i$  linear forms of the recursive type

$$(52) \quad \zeta_i = \xi_i + \sum_{(j>i)} \xi_j \beta_{ji}.$$

The product  $q_1 \dots q_n = D = D_n$  is called the discriminant of  $f$ .

Break the sum (51) into two parts according to

$$f(\mathfrak{x}) = (q_1 |\zeta_1|^2 + \dots + q_{k-1} |\zeta_{k-1}|^2) + (q_k |\zeta_k|^2 + \dots + q_n |\zeta_n|^2)$$

and substitute  $\mathfrak{x} = \mathfrak{e}_k$ . The value of the whole form is  $\gamma_{kk}$  while the value of the second summand is  $q_k$ . Hence

$$(53) \quad q_k \leq \gamma_{kk},$$

$$(54) \quad D \leq \gamma_{11} \dots \gamma_{nn}.$$

The Jacobi transformation of the positive form

$$f^{(k)} = f(\xi_1, \dots, \xi_k, 0, \dots, 0)$$

of  $k$  variables is obtained from (51) by setting  $\xi_{k+1} = \dots = \xi_n = 0$ . Consequently its discriminant is  $D_k = q_1 \dots q_k$  and thus

$$q_k = D_k / D_{k-1} \quad (k = 1, \dots, n; D_0 = 1).$$

The first step (49) goes through under the sole assumption  $\gamma_{11} = q_1 > 0$ . If, in carrying the process further for a given form  $f$ , we find  $q_2 > 0, \dots, q_n > 0$

at the following steps, then the formula (51) itself reveals that  $f$  is positive.

By (50) the inequality  $q_2 > 0$  amounts to

$$|\gamma_{21}|^2 = |\gamma_{12}|^2 < \gamma_{11} \cdot \gamma_{22}.$$

More generally we must have

$$|\gamma_{ij}|^2 < \gamma_{ii} \cdot \gamma_{jj} \quad (i \neq j)$$

for any positive form  $f$ .

Next we compute the volume  $V$  of the  $4n$ -dimensional ellipsoid  $f(x) < 1$ . Denote by  $\omega_n$  the volume of the sphere

$$x_1^2 + \cdots + x_n^2 < 1$$

in the  $n$ -dimensional real vector space. When in the recursive substitution (52) we replace each of the quaternions  $\xi$  and  $\zeta$  by its 4 real  $x$ -components, we again obtain a recursive substitution, this time in  $4n$  variables, whose coefficient matrix has 1's along the principal diagonal and hence is of determinant 1. Thus the volume  $V$  is the same as that of the Jacobi transform  $\sum q_i |\xi_i|^2 < 1$  or in real  $x$ -components

$$\sum_{i=1}^n \sum_{\alpha=0}^3 (q_i^{1/2} x_{i\alpha})^2 < 1.$$

Consequently

$$(55) \quad V = \frac{\omega_{4n}}{(q_1^{1/2} \cdots q_n^{1/2})^4} = \frac{\pi^{2n}}{(2n)!} \frac{1}{q_1^2 \cdots q_n^2} = \frac{\pi^{2n}}{(2n)!} \frac{1}{D^2}.$$

In the real and the imaginary case one finds

$$(56) \quad V = \frac{\omega_n}{D^{1/2}}, \quad V = \frac{\omega_{2n}}{D} = \frac{\pi^n}{n!} \frac{1}{D}$$

instead. Incidentally these formulas prove that, although our recursive definition refers to a definite arrangement, the discriminant of  $f$  is not changed by arranging the variables  $\xi_1, \dots, \xi_n$  in a different order.

From here on we limit ourselves to real quadratic forms, because the adjustments to the two other cases are sufficiently trivial; only an occasional glance will be cast upon them.

**9. Some simple topological considerations.** Within the  $N$ -dimensional linear space  $R$  of all quadratic forms  $f = \{g_{ij}\}$  the positive ones form a convex subset  $G$  which is a cone with the origin  $f=0$  as vertex. The relative clause means that dilatation,  $f \rightarrow tf$ , at any positive rate  $t$  carries  $G$  into itself.  $G$  is an open set. Indeed the quantities emerging at the first step of Jacobi's transformation,

$$q_1 = g_{11}, \quad b_{i1} = \frac{g_{i1}}{g_{11}}, \quad g_{ij}^* = g_{ij} - \frac{g_{i1}g_{1j}}{g_{11}} \quad (i, j = 2, \dots, n),$$

all depend continuously on  $f$ . [We now use corresponding Roman instead of Greek letters throughout, so that the transformation (52) reads

$$(52') \quad z_i = x_i + \sum_{(j>i)} x_j b_{ji}.]$$

Hence  $q_1, \dots, q_n; b_{ji} (j>i)$  depend continuously on  $f$  at a given point  $f^0$  of  $G$ , and all forms  $f$  in a certain neighborhood  $U$  of  $f^0$  will satisfy the conditions

$$q_1 \geq \frac{1}{2}q_1^0, \dots, q_n \geq \frac{1}{2}q_n^0$$

and thus be positive.

Jacobi's transformation shows quite explicitly that for a given positive form  $f$  and a given number  $A$  the inequality  $f(\mathfrak{x}) \leq A$  entails upper bounds for the  $|x_i|$  of  $\mathfrak{x} = (x_1, \dots, x_n)$ . In fact, one first obtains upper bounds for  $|z_1|, \dots, |z_n|$  and then, going in backward direction, from the relations (52') upper bounds for  $|x_n|, |x_{n-1}|, \dots, |x_1|$ . One can make this estimate uniform throughout a sufficiently small neighborhood of a given form. Hence this

**LEMMA 4.** *Let  $A(f)$  be a real function depending on a variable point  $f$  in  $G$  and continuous at the given point  $f^0$ . We can fix a neighborhood  $U$  of  $f^0$  such that nearly every lattice vector  $\mathfrak{x} = (x_1, \dots, x_n)$  has the property of satisfying the inequality*

$$f(\mathfrak{x}) > A(f)$$

for all  $f$  in  $U$ .

("Nearly every" means that only a finite number lack the property in question.)

**Proof.** We fix the neighborhood  $U$  so that

$$q_k(f) \geq \frac{1}{2}q_k^0, \quad A(f) \leq 1 + A(f^0), \quad |b_{ji}(f)| \leq 1 + |b_{ji}^0|.$$

If  $\mathfrak{x}$  is a vector such that there is an  $f$  in  $U$  for which  $f(\mathfrak{x}) \leq A(f)$ , then (51) yields upper bounds for  $|z_k|$  which are universal in that they do not depend on the specific  $f$  in  $U$ , and (52') yields universal bounds for  $|x_n|, \dots, |x_1|$ .

From now on up to the end of §12,  $f$  without or with accent or index always indicates a point of  $G$ . All topological notions are to be interpreted relative to  $G$ ; e.g., a subset of  $G$  is said to be open or closed whenever it is open or closed relative to  $G$ .

Before going on we specialize some of our previous definitions concerning gauge functions to gauge functions of the type  $f^{1/2}$  now under consideration. A positive quadratic form  $f$  is said to be reduced if it satisfies the inequality

$$f(x_1, \dots, x_n) \geq g_{kk}$$

for any vector  $(x_1, \dots, x_n)$  in  $X_k$  and for  $k=1, \dots, n$ . This implies

$$(0 <) g_{11} \leq g_{22} \leq \dots \leq g_{nn}.$$

Two forms  $f, f'$  are called equivalent and counted in the same class if one proceeds from the other by a substitution

$$x_i = \sum_k x'_k s_i^k$$

of the modular group. Every point  $f$  in  $G$  is equivalent to a reduced one.

To each index  $k$  and vector  $\mathfrak{x} = (x_1, \dots, x_n)$  in  $X_k$  there corresponds a linear form of the coordinates  $g_{ij}$  in  $R$ ,

$$f(\mathfrak{x}) - g_{kk} = \sum_{i,j} x_i x_j g_{ij} - g_{kk} = \sum_{i,j} \alpha_{ij} g_{ij},$$

which we denote by  $\alpha_k(\mathfrak{x})$ ; its coefficients are

$$\alpha_{ij} = x_i x_j - \delta_i^k \delta_j^k.$$

The relations for the variable point  $\{g_{ij}\}$ ,

$$\sum \alpha_{ij} g_{ij} = 0, \quad \geq 0, \quad > 0$$

are referred to as the *equation*, the *inequality* and the *strict inequality*  $\alpha_k(\mathfrak{x})$  respectively. Except for  $\mathfrak{x} = \pm e_k$ , i.e., for every vector  $\mathfrak{x}$  in  $X_k^*$ , the inequality and equation  $\alpha_k(\mathfrak{x})$  define a half-space and its bounding  $(N-1)$ -dimensional plane in  $R$ . Now  $f$  is properly reduced provided the strict inequality  $\alpha_k(\mathfrak{x})$  is satisfied for every  $\mathfrak{x}$  in  $X_k^*$  and every  $k$ . Examples of properly reduced forms are ready at hand; the simplest are the diagonal forms

$$g_1 x_1^2 + \dots + g_n x_n^2 \quad \text{with } 0 < g_1 < g_2 < \dots < g_n.$$

The reduced points form a closed convex subset  $Z$  of  $G$  which again is a cone and will be called the (basic) *cell*. A properly reduced  $f$  is said to belong to the *core* of  $Z$ . An inner point of  $Z$  belongs to its core. Each unimodular substitution  $S$  carries  $Z$  into an equivalent cell  $Z_S$ . The substitutions of the subgroup  $\{J\}$  leave  $Z$  unchanged, but if  $S$  is not in  $\{J\}$  then no point of the core of  $Z$  can be in  $Z_S$  (Theorem 7). Hence the equivalent cells  $Z_S$  cover  $G$  without gaps and overlappings; two different cells have none but boundary points in common. Here two substitutions like  $S$  and  $JS$  which are left equivalent modulo  $\{J\}$  are to be identified because they have the same effect on  $Z$ . Our aim is first to study the individual cell  $Z$  and then the whole pattern of the division of  $G$  into equivalent cells.

We start with the observation that a point  $f^0$  belonging to the core of  $Z$  is

an inner point of  $Z$ . Indeed according to Lemma 4, nearly every lattice vector  $\mathfrak{x}$  satisfies the inequalities  $f(\mathfrak{x}) > g_{kk}$  for  $k=1, \dots, n$  and for all forms  $f$  in a certain neighborhood  $U$  of  $f^0$ . Therefore among the infinitely many inequalities

$$(57) \quad \alpha_k(\mathfrak{x}) \quad (\mathfrak{x} \text{ in } X_k^*; k=1, \dots, n)$$

there are only a finite number, say  $\alpha', \alpha'', \dots$ , which are not a priori sure to hold throughout  $U$ . But if the strict inequalities  $\alpha', \alpha'', \dots$  hold for  $f^0$  then they hold also in a sufficiently small neighborhood  $U'$  of  $f^0$ ; and the neighborhood  $U \cap U'$  of  $f^0$  lies in  $Z$ .

Denote by  $T_k$  the subset of  $X_k$  to which  $\mathfrak{x}$  belongs if there are reduced forms  $f$  satisfying the equation  $f(\mathfrak{x}) = g_{kk}$ . The two vectors  $\pm \mathfrak{e}_k$  belong to  $T_k$ , and again  $T_k^*$  designates what is left of  $T_k$  after these two vectors have been removed. The planes  $\alpha_k(\mathfrak{x}) = 0$  corresponding to the  $\mathfrak{x}$  in  $T_k^*$  graze the cell  $Z$ . Our last result asserts that every boundary point of  $Z$  lies in one of these grazing planes

$$(58) \quad \alpha_k(\mathfrak{x}) = 0 \quad (\mathfrak{x} \text{ in } T_k^*, k=1, \dots, n).$$

Hence from a general topological principle which we shall presently prove for our special situation there follows

**THEOREM 10.** *In the definition of  $Z$  as the set consisting of all points  $f$  of  $G$  which satisfy the inequalities*

$$(59) \quad \alpha_k(\mathfrak{x}) \quad \text{for every } \mathfrak{x} \text{ in } X_k \text{ and } k=1, \dots, n,$$

*the vector set  $X_k$  may be replaced by  $T_k^*$ .*

**Proof.** Choose one of the points  $f^0$  belonging to the core of  $Z$  as the center of  $Z$  and suppose  $f$  is any point (of  $G$ ) outside  $Z$ . Join  $f^0$  with  $f$  by a straight segment. Somewhere, at a point  $f'$ , it will cross the border of  $Z$ ; the part  $f^0 f'$  of the segment, including  $f'$ , belongs to  $Z$  while the points beyond  $f'$  are outside  $Z$ . The point  $f'$  satisfies one of the equations (58), say

$$(60) \quad \sum \alpha_{ij} g_{ij} = 0.$$

The left member of (60) is greater than 0 at  $f^0$ , equals 0 at  $f'$ , and hence is less than 0 at  $f$ . Consequently a point  $f$  which satisfies all inequalities

$$\alpha_k(\mathfrak{x}) \geq 0 \quad (\mathfrak{x} \text{ in } T_k^*, k=1, \dots, n)$$

cannot lie outside  $Z$  [14].

We denote by  $X_k^0$  the set of lattice vectors  $(x_1, \dots, x_n)$  for which

$$x_k = 1, \quad x_{k+1} = \dots = x_n = 0.$$

$X_k^0$  is a subset of  $X_k$ . Let  $\rho$  be any number greater than 1 and  $\sigma$  a positive

number. Later on we shall have occasion to study the part  $G(\rho, \sigma)$  of  $G$  defined by the following simultaneous inequalities:

$$(61_1) \quad f(x_1, \dots, x_n) \geq \frac{1}{\rho^2} g_{kk} \quad \text{for every vector } (x_1, \dots, x_n) \text{ in } X_k,$$

$$(61_2) \quad f(x_1, \dots, x_n) \geq g_{kk} - \sigma g_{11} \quad \text{for every vector } (x_1, \dots, x_n) \text{ in } X_k^0 \\ [k = 1, \dots, n].$$

$G(\rho, \sigma)$  is a closed convex part of  $G$  which increases with increasing  $\rho$  and  $\sigma$ . A point  $f$  of  $G$  satisfying all these inequalities (61) with the  $>$  sign is an inner point of  $G(\rho, \sigma)$ , as follows by the argument previously applied to  $Z$ . The domain  $G(\rho, \sigma)$  contains the cell  $Z$  in its interior. I propose to show that with  $\rho \uparrow \infty, \sigma \uparrow \infty$  it exhausts the whole  $G$ . Let  $f$  be any point of  $G$ . All lattice vectors  $(x_1, \dots, x_n)$  except those of a certain finite set  $\Sigma$  satisfy the inequalities

$$f(x_1, \dots, x_n) > g_{kk} \quad (k = 1, \dots, n)$$

and hence (61), whatever the values  $\rho > 1$  and  $\sigma > 0$ . When  $(x_1, \dots, x_n)$  varies over the finite set  $X_k \cap \Sigma$ ,  $f(x_1, \dots, x_n)$  will assume a least (positive) value  $g_{kk}/\rho_k^2$ . Thus all the inequalities (61), with the  $>$  sign and for  $k = 1, \dots, n$ , will hold as soon as  $\rho > \rho_1, \rho_2, \dots, \rho_n$ . In the same manner one sees that, for a sufficiently high  $\sigma$ ,  $f$  satisfies all relations (61<sub>2</sub>) with the  $>$  sign for  $k = 1, \dots, n$ .

**10. The first theorem of finiteness.** We now resume the algebraic study of reduced forms, first specializing Theorem 5 for the gauge function  $f^{1/2}$ :

**THEOREM 11.** *Any reduced form  $f = \{g_{ij}\}$  satisfies the inequality*

$$(62) \quad \lambda_n g_{11} \cdots g_{nn} \leq D$$

where  $\lambda_n = (\omega_n/\mu_n)^2$ .

About the constant  $\mu_n$  see the Supplement to Theorems 4-6 in §3. We use the formulas (55) and (56) for the volumes of our ellipsoidal gauge bodies and thus obtain a corresponding inequality

$$\lambda_n \gamma_{11} \cdots \gamma_{nn} \leq D$$

for reduced Hermitian and Hamiltonian forms, with

$$\lambda_n = \frac{\pi^n}{n!} \cdot \frac{1}{\mu_n^2}, \quad \lambda_n = \frac{\pi^n}{[(2n)!]^{1/2}} \cdot \frac{1}{\mu_n^2}$$

and the values of  $\mu_n^2$  given by Theorem 5\*\*. The resulting values of  $\lambda_n$  are certainly not optimal, but fairly good.

In passing we mention the following relations:



$$(63) \quad |g_{ij}| \leq \frac{1}{2}g_{ii}, \quad |\gamma_{ij}| \leq r\gamma_{ii},$$

which hold for reduced forms and for  $i < j$ . Choose two different indices, say 2 and 5. The two vectors for which  $x_2 = \pm 1$ ,  $x_5 = 1$  and all other  $x_i$  vanish belong to  $X_5$ ; hence

$$g_{22} \pm 2g_{25} + g_{55} \geq g_{55}$$

or

$$2|g_{25}| \leq g_{22}.$$

In the imaginary and quaternion cases the procedure is as follows. Let  $\eta$  range over all integers. We take  $\xi_5 = 1$  and  $\xi_2 = -\eta$  while all other  $\xi_i$  vanish. The resulting inequality reads

$$\gamma_{22}\eta\bar{\eta} - \eta\gamma_{25} - \gamma_{52}\bar{\eta} \geq 0$$

which for  $\gamma = \gamma_{52}/\gamma_{22}$  yields

$$|\gamma - \eta|^2 \geq |\gamma|^2.$$

This means that, in the lattice of integers,  $\gamma$  is not farther from zero than from any other integer. Hence this distance  $|\gamma|$  cannot exceed  $r$ .

If  $f(x_1, \dots, x_n)$  is a reduced form of  $n$  variables, then

$$f^{(k)} = f(x_1, \dots, x_k, 0, \dots, 0)$$

is one of  $k$  variables, therefore

$$D_k \geq \lambda_k g_{11} \cdots g_{kk}.$$

Combining this with (54) for  $f^{(k-1)}$ ,  $D_{k-1} \leq g_{11} \cdots g_{k-1, k-1}$ , we find the important inequality

$$q_k \geq \lambda_k g_{kk} \quad (k = 1, \dots, n)$$

holding for reduced forms  $f$ .

We are now sufficiently prepared to prove the first theorem of finiteness:

**THEOREM 12.** *The set  $T_k$  of lattice vectors is finite.*

Hence by Theorem 10 we have succeeded in sifting from the infinitely many inequalities (59) a finite number on which all others are consequent and therefore redundant. In proving our proposition we shall give fairly explicit upper bounds of  $|x_1|, \dots, |x_n|$  for the vectors  $\mathfrak{x} = (x_1, x_2, \dots, x_n)$  in  $T_k$ .

**Proof.** Suppose  $\mathfrak{x}$  is in  $T_k$  and  $f$  a reduced form for which  $f(\mathfrak{x}) = g_{kk}$ . In particular  $\mathfrak{x} = \mathfrak{e}_k$  fulfills this demand. We apply Jacobi's transformation to  $f$  and then find for the vector in question

$$q_1 z_1^2 + \cdots + q_n z_n^2 = g_{kk};$$

a fortiori

$$\sum_{j=k}^n q_j z_j^2 \leq g_{kk}.$$

In the last sum  $q_j \geq \lambda_j g_{jj} \geq \lambda_j g_{kk}$  ( $j \geq k$ ), and thus the inequality

$$\sum_{j=k}^n \lambda_j z_j^2 \leq 1$$

results which yields universal upper bounds for  $|z_k|, \dots, |z_n|$ :

$$(64) \quad |z_j|^2 \leq 1/\lambda_j \quad (j = k, \dots, n).$$

To find universal bounds for  $|z_1|, \dots, |z_{k-1}|$  is a slightly more intricate job. Let  $h$  be a given index less than  $k$ . Without altering  $x_n, \dots, x_{h+1}$  we may replace  $x_h, \dots, x_1$  by such integers  $x_h^*, \dots, x_1^*$  in succession that the corresponding  $z_h^*, \dots, z_1^*$  satisfy

$$|z_h^*| \leq \frac{1}{2}, \dots, |z_1^*| \leq \frac{1}{2}.$$

Since the new vector  $(x_1^*, \dots, x_h^*, x_{h+1}, \dots, x_n)$  also is in  $X_k$ , we must have

$$f(x_1^*, \dots, x_h^*, x_{h+1}, \dots, x_n) \geq g_{kk},$$

consequently

$$\begin{aligned} (q_1 z_1^{*2} + \dots + q_h z_h^{*2}) + (q_{h+1} z_{h+1}^2 + \dots + q_n z_n^2) \\ \geq g_{kk} = (q_1 z_1^2 + \dots + q_h z_h^2) + (q_{h+1} z_{h+1}^2 + \dots + q_n z_n^2) \end{aligned}$$

or

$$q_1 z_1^{*2} + \dots + q_h z_h^{*2} \geq q_1 z_1^2 + \dots + q_h z_h^2.$$

The left member is less than or equal to

$$r^2(q_1 + \dots + q_h) \leq r^2(g_{11} + \dots + g_{hh}) \leq r^2 h g_{hh} \quad (r = \frac{1}{2}).$$

Hence

$$r^2 h g_{hh} \geq q_h z_h^{*2} \geq \lambda_h g_{hh} z_h^{*2} \quad \text{or}$$

$$(65) \quad z_h^2 \leq r^2 h / \lambda_h \quad (h = 1, \dots, k-1).$$

(The notation  $r$  is used in order to cover also the imaginary and quaternion cases.)

Applying (65) to  $z = e_k$ , one gets

$$(66) \quad b_{kh}^2 \leq r^2 h / \lambda_h \quad (\text{for } h < k).$$

The universal upper bounds for  $|z_n|, \dots, |z_1|$  together with the universal bounds for the moduli of the coefficients  $b_{kh}$  in the recursive equations (52') result in universal upper bounds for  $|x_n|, \dots, |x_1|$ .

This argument is chiefly due to Minkowski and is in my view the backbone of his theory of reduction. The simple remark leading from (65) to (66) was first made by Remak [15]. It dispenses with the necessity of making use of the explicit expression of  $b_{kh}$  as Minkowski did, which is the more fortunate as it would have been quite cumbersome to follow his procedure in the quaternion case.

**11. The second theorem of finiteness. Generators of the modular group.** We prove now the following theorem.

**THEOREM 13.** *The set  $G(\rho, \sigma)$  has points in common with not more than a finite number of cells  $Z_s$ .*

We must show that there is only a finite number of unimodular substitutions  $S$  capable of carrying an (unspecified) point  $f$  of  $Z$  into a point  $f'$  of  $G(\rho, \sigma)$ .

$$f(y_1\delta_1 + \cdots + y_n\delta_n) = f'(y_1, \cdots, y_n).$$

Here  $(\delta_1, \cdots, \delta_n)$  is a lattice basis of the property  $B(\rho, \cdots, \rho)$  with respect to  $f^{1/2}$ . Consider the two series of subspaces

$$E_0, E_1 = [e_1], E_2 = [e_1, e_2], \cdots, E_n;$$

$$E'_0, E'_1 = [\delta_1], E'_2 = [\delta_1, \delta_2], \cdots, E'_n.$$

$E_0 = E'_0$  is the zero space,  $E_n = E'_n$  the full vector space. Let  $l$  be the highest of the indices  $1, \cdots, n$  for which

$$(67) \quad E'_{l-1} = E_{l-1}.$$

The decision whether or not  $E'_k = E_k$  depends merely on checking whether some integers are zero. For  $l$  there exist the possibilities  $l=1, \cdots, n$ . We propose to consider the  $S$  with a definite  $l$ .

First we focus our attention on the vectors

$$(68) \quad \delta_k \quad (k = l, \cdots, n).$$

For the moment let  $\mathbf{x} = (x_1, \cdots, x_n)$  denote the vector  $\delta_k$ ; then

$$(69) \quad f(\mathbf{x}) = q_1 x_1^2 + \cdots + q_n x_n^2 = g'_{kh}.$$

Put

$$\theta'_i = \theta_i(p) \quad \text{for } p_1 = \cdots = p_i = \rho.$$

Because of the significance of  $l$  and Theorem 9<sub>p</sub> we have

$$(\theta'_i \theta_{i+1})^2 g_{ii} \geq g_{i+1, i+1}$$

for  $i \geq l$ . Therefore and because  $f$  is reduced,

$$\begin{aligned}
 q_1 z_1^2 + \cdots + q_n z_n^2 &\geq \lambda_1 g_{11} z_1^2 + \cdots + \lambda_n g_{nn} z_n^2 \\
 (70) \qquad &\geq g_{kk} \left\{ \lambda_n z_n^2 + \cdots + \lambda_k z_k^2 + \frac{\lambda_{k-1}}{(\theta'_{k-1} \theta_k)^2} z_{k-1}^2 + \cdots \right. \\
 &\quad \left. + \frac{\lambda_1}{(\theta'_{k-1} \cdots \theta'_1 \cdot \theta_k \cdots \theta_{l+1})^2} z_1^2 \right\},
 \end{aligned}$$

while by Theorem 8<sub>p</sub>,

$$(71) \qquad g'_{kk} \leq (\theta'_k)^2 g_{kk}.$$

Combining the two inequalities (70) and (71) with (69), we get hold of universal upper bounds for  $|z_l|, \dots, |z_n|$ , namely

$$\begin{aligned}
 |z_k| &\leq \frac{\theta'_k}{\lambda_k^{1/2}}, \dots, |z_n| \leq \frac{\theta'_n}{\lambda_n^{1/2}}, \\
 |z_{k-1}| &\leq \frac{\theta'_{k-1} \theta'_k \cdots \theta'_1 \cdot \theta_k}{\lambda_{k-1}^{1/2}}, \dots, |z_l| \leq \frac{\theta'_l \cdots \theta'_k \cdot \theta_{l+1} \cdots \theta_k}{\lambda_l^{1/2}}.
 \end{aligned}$$

So far we have used merely the first set (61<sub>1</sub>) of inequalities for  $f'$ .

The second set yields universal bounds for  $|z_1|, \dots, |z_{l-1}|$ . Suppose  $y_1, \dots, y_{l-1}$  to be any integers; we have

$$f'(y_1, \dots, y_{l-1}, \delta_l^k, \dots, \delta_n^k) \geq f'(\delta_1^k, \dots, \delta_n^k) - \sigma'_{g_{11}}$$

which is equivalent to

$$f(y_1 \delta_1 + \cdots + y_{l-1} \delta_{l-1} + \delta_k) \geq f(\delta_k) - \sigma'_{g_{11}}$$

or

$$(72) \qquad f(x_1^*, \dots, x_{l-1}^*, x_l, \dots, x_n) \geq f(x_1, \dots, x_n) - \sigma'_{g_{11}}$$

where  $(x_1, \dots, x_n)$  again is the vector  $\delta_k$  and  $x_1^*, \dots, x_{l-1}^*$  denote any integers. In fact

$$x_i^* = x_i + x'_i, \dots, x_{l-1}^* = x_{l-1} + x'_{l-1}$$

with

$$y_1 \delta_1 + \cdots + y_{l-1} \delta_{l-1} = x'_1 \epsilon_1 + \cdots + x'_{l-1} \epsilon_{l-1}.$$

Observe that  $\delta_1, \dots, \delta_{l-1}$  span the lattice in  $E_{l-1}$  so that  $x'_1, \dots, x'_{l-1}$  and therefore  $x_1^*, \dots, x_{l-1}^*$  range independently over all integers while  $y_1, \dots, y_{l-1}$  do so. Let  $h$  be one of the indices  $1, \dots, l-1$ , and choose  $x_i^* = x_i$  for  $i > h$ , but  $x_h^*, \dots, x_1^*$  such that

$$|z_h^*| \leq r, \dots, |z_l^*| \leq r \quad (r = \frac{1}{2}).$$

Then (72) yields

$$\sigma g'_{11} + (q_1 z_1^{*2} + \dots + q_h z_h^{*2}) \geq q_1 z_1^2 + \dots + q_h z_h^2 \geq q_h z_h^2 \geq \lambda_h g_{hh} z_h^2.$$

The left member is less than or equal to

$$\sigma \rho^2 g_{11} + r^2 (g_{11} + \dots + g_{hh}) \leq (\sigma \rho^2 + r^2 h) g_{hh};$$

thus

$$\lambda_h z_h^2 \leq \sigma \rho^2 + r^2 h \quad (h = 1, \dots, l-1).$$

Hence we have obtained universal bounds for all  $|z_i|$  and by means of (66) also for all  $|x_i|$ . In other words, for each of the lattice vectors (68) we find ourselves limited to a finite set from which to choose.

If  $l=1$  nothing remains to be said. In the opposite case the same situation prevails for the "cut" forms

$$f(x_1, \dots, x_{l-1}, 0, \dots, 0), \quad f'(x_1, \dots, x_{l-1}, 0, \dots, 0)$$

of  $l-1 < n$  variables in  $E_{l-1}$  as for the full forms  $f$  and  $f'$  in  $n$  dimensions which we started with. Thus the proof is complete by induction.

The main idea of the proof is again borrowed from Minkowski—with two essential modifications:

(1) Where Minkowski uses estimates based upon Jacobi's transformation of quadratic forms, we have availed ourselves of the general Theorems 8 and 9 holding for any gauge function whatsoever; in spite of their far greater generality these estimates are sharper than Minkowski's.

(2) Minkowski has our proposition only for

$$\rho = 1, \quad \sigma = 0, \quad G(\rho, \sigma) = Z,$$

in which case it asserts that  $Z$  borders on not more than a finite number of equivalent cells  $Z_s$ . However, we should know that every boundary point of  $Z$  is on the common boundary of  $Z$  and a different cell  $Z_s$ , or that the cells  $Z_s$  cluster only towards the border of  $G$ , which means that into any sufficiently small neighborhood of a point of  $G$ , or into any compact subset of  $G$ , there penetrate only a finite number of cells  $Z_s$ . Our theorem goes beyond this because  $G(\rho, \sigma)$  exhausts  $G$  if  $\rho \uparrow \infty, \sigma \uparrow \infty$ , but is not compact. About this finer point refer to §13. Here is an application of the fact that the cells do not cluster in the interior of  $G$ :

LEMMA 5. Any cell  $Z' = Z_s$  may be reached from the basic cell  $Z$  by a chain

$$(73) \quad Z = Z_1, Z_2, \dots, Z_r = Z'$$

in which any two consecutive members are in contact, i.e., have points in common.

**Proof.** The center  $f^0$  of  $Z$  goes by the substitution  $S$  into an inner point  $f_s^0$  of  $Z_s = Z'$ . Join  $f^0$  with  $f_s^0$  by a straight segment  $\tau$ . Determine  $\rho > 1$  and  $\sigma > 0$  so that  $f_s^0$  is an inner point of  $G(\rho, \sigma)$ . Then the whole segment  $\tau$  lies in  $G(\rho, \sigma)$ . Since the number of cells  $Z_s$  having points in common with  $G(\rho, \sigma)$  is finite, the same is true a fortiori for the cells  $Z_s$  which are met by the segment  $\tau$ . On the other hand every point of  $\tau$  belongs to a certain cell  $Z_s$ , and the points which  $\tau$  and  $Z_s$  have in common form a (closed) interval on  $\tau$ . Hence  $\tau$  is covered by a finite number of subintervals of which we can select a chain connecting  $f^0$  with  $f_s^0$ . What we obtain in this manner is a chain of cells (73) in which any two consecutive members have a contact point on  $\tau$ .

Those substitutions of the modular group which effect transition from  $Z$  to cells in contact with  $Z$  form a finite set  $[Z]$ . If  $S$  is in  $[Z]$ , so is the "(two-sided) congruent" substitution

$$S^* = JSJ' \quad (J, J' \text{ any two elements of } \{J\})$$

as one readily verifies by performing the substitution  $J'$  on the two contacting cells  $Z$  and  $Z_{JS} = Z_s$ . Hence  $[Z]$  breaks up into a number of complete sets of congruent substitutions; we choose a representative out of each set:  $S', S'', \dots$ .

**THEOREM 14.** *The substitutions of  $\{S\}$  which carry  $Z$  into cells bordering on  $Z$ , or rather a complete system of modulo  $\{J\}$  incongruent representatives  $S', S'', \dots$  among them, combined with  $\{J\}$ , generate the whole modular group  $\{S\}$ .*

**Proof.** Let  $S$  be any element of the modular group and determine a chain (73) leading from  $Z$  to  $Z_s = Z'$ . A certain unimodular  $S_i^{-1}$  will carry  $Z_i$  into  $Z$  and  $Z_{i+1}$  into a cell contacting  $Z$  which therefore arises from  $Z$  by an element  $S^{(i)}$  of  $[Z]$ . The substitution  $S^{(i)}S_i$  carries  $Z$  into  $Z_{i+1}$  and thus can and shall be adopted as  $S_{i+1}$ . If this inductive definition of  $S_i$  is started off with  $S_1$  the identity, then  $S^{(r-1)} \dots S^{(1)}$  carries  $Z$  into  $Z_s$ , and therefore

$$S = JS^{(r-1)} \dots S^{(1)} \quad (J \text{ in } \{J\}).$$

**12. Faces and walls.** The main body of the theory of reduction is now complete; what follows are accessories of minor importance. In this section we discuss the consequences upon the cell configuration of the fact that any boundary point of a convex solid polyhedron lies on one of its faces. Engaging in this kind of general topological argument, we prefer the notation  $y_1, \dots, y_N$  instead of  $g_{ij}$  for the coordinates in our  $N$ -dimensional space  $R$ . A face of the cell would be described by one of the equations

$$(57) \quad \alpha_k(x) \quad (x \text{ in } X_k^*, k = 1, \dots, n),$$

which hold for  $N-1$  linearly independent points of  $Z$ . Taking it for granted that each boundary point of  $Z$  lies on a face, we infer from the proof of



Theorem 10 that the corresponding inequalities suffice to define  $Z$  as a part of  $G$ : those planes (57) which do not share an  $(N-1)$ -dimensional convex face with  $Z$  may be discarded. It is clear that on account of their "extreme" character the remaining inequalities are truly indispensable.

As to the configuration of all equivalent cells  $Z_s$ , it seems clear that any point on the boundary of  $Z$  lies on a "wall" separating  $Z$  from an "adjacent" cell  $Z_s$ . By these words "wall" and "adjacent" we wish to indicate that  $Z$  and  $Z_s$  have  $N-1$  linearly independent points in common. The points which two cells have in common, if any, form a convex cone of 1 or 2 or  $\dots$  or  $N-1$  dimensions. We speak of a contact of order 1, 2,  $\dots$ ,  $N-1$  respectively. The unimodular  $S$  carrying  $Z$  into adjacent cells form a finite set  $[[Z]]$  narrower than  $[Z]$ . Again it decomposes into subsets of congruent substitutions. Theorem 14 remains true if  $S', S'', \dots$  denote representatives of these sets. We can dispense with none of these more restricted generators.

The ultimate goal of all such considerations should be to show that the pattern of our cells which mutually border on each other is a *complex* in the combinatorial topological sense, of such particular structure as to form the skeleton of a manifold.

It is clear that the walls of  $Z$  are parts of its faces. This simple observation establishes a close relationship between the first and Minkowski's special case of the second theorem of finiteness.

I shall try to give the most convenient arrangement of the proofs. First the faces of  $Z$ .

LEMMA 6. *Any boundary point of  $Z$  lies on a face of  $Z$ .*

We know that  $Z$  as a part of  $G$  is characterized by inequalities

$$(74) \quad \alpha(y) = \alpha_1 y_1 + \dots + \alpha_N y_N \geq 0$$

corresponding to a finite set  $\Sigma = \Sigma_0$  of linear forms  $\alpha(y)$ . Let  $f^1$  be a point (of  $G$ ) on the boundary of  $Z$ ; it will satisfy at least one of the inequalities of  $\Sigma$  with the  $=$  sign. After an appropriate linear transformation of the coordinates  $y$ , we may assume

$$(75) \quad f^1 = e^1 = (1, 0, 0, \dots, 0).$$

$\Sigma_1$  is the non-empty subset of  $\Sigma$  to which a linear form  $\alpha(y)$  belongs if nullified by  $e^1$ . Their first coefficient  $\alpha_1$  vanishes, so that they may be looked upon as forms of  $N-1$  variables. For the linear forms  $\alpha(y)$  in the complementary subset  $\bar{\Sigma}_0$  the first coefficient  $\alpha_1$  is positive. We describe the  $\nu$ th step of this process of selection. Suppose the subset  $\Sigma_\nu$  of those linear forms of  $\Sigma$  in which the variables  $y_1, \dots, y_\nu$  are absent is not empty. The corresponding inequalities

$$\alpha_{\nu+1} y_{\nu+1} + \dots + \alpha_N y_N \geq 0$$

of  $\Sigma_\nu$  define a convex pyramid  $Z'$  in the  $(N-\nu)$ -dimensional space  $R'$  with

the coordinates  $y_{r+1}, \dots, y_N$ . As long as  $N - \nu \geq 2$ , we can find a point  $f^{r+1} \neq 0$  on the boundary of that pyramid, and by a suitable affine transformation of the coordinates  $y_{r+1}, \dots, y_N$  we can provide for  $f^{r+1}$  having the coordinates

$$(y_{r+1}, \dots, y_N) = (1, 0, \dots, 0).$$

$\Sigma_r$  breaks up into the subsets  $\Sigma_{r+1}$  and  $\bar{\Sigma}_r$ , whose members have their first coefficient  $\alpha_{r+1} = 0$  and  $> 0$  respectively.  $\Sigma_{r+1}$  is not empty.

The existence of  $f^{r+1}$  follows in this way. Denote by  $f^0 = (y_1^0, \dots, y_N^0)$  the center of the cell  $Z$ . All linear forms  $\alpha(y)$  belonging to  $\Sigma_r$  have the property  $\alpha(f^0) > 0$  for

$$(76) \quad f^0 = (y_{r+1}^0, \dots, y_N^0),$$

or (76) is an inner point of  $Z^r$ . Operating in the  $(N - \nu)$ -dimensional space  $R^r$  we choose one of the forms of  $\Sigma_r$ , say  $\alpha'(y)$ , and a point  $f \neq 0$  in the plane  $\alpha'(y)$ . (As long as  $R^r$  has at least two dimensions, a plane  $\alpha'(y) = 0$  through the origin  $O$  certainly contains points  $f \neq 0$ .) We join  $f^0$  with  $f$  by a straight segment, which will not contain the origin  $O$ . Traveling along the segment from  $f^0$  to  $f$  we encounter a first point  $f^*$  where one of the forms of  $\Sigma_r$  ceases to be positive. (If not before this will happen for  $f$ .) All forms of  $\Sigma_r$  are greater than or equal to 0 for  $f^*$  and at least one equals 0. We take  $f^{r+1} = f^*$ .

We end up with a non-empty set  $\Sigma_{N-1}$  consisting of linear forms  $\alpha_N y_N$  in the 1-dimensional space  $R^{N-1}$  with the single coordinate  $y_N$ . They are positive for  $y_N = y_N^0$ . We take one of them as the coordinate  $y_N$ ; then the coefficients  $\alpha_N$  of the others are greater than 0 and  $y_N \geq 0$  is the pyramid  $Z^{N-1}$  in  $R^{N-1}$ . At the same time we have arrived at a complete normalization of the affine system of coordinates  $y_1, \dots, y_N$ .

By construction the pyramid  $Z^{r-1}$  in  $R^{r-1}$  contains the point

$$(y_r, \dots, y_N) = (1, 0, \dots, 0).$$

The system  $\Sigma_{r-1}$  of linear forms

$$\alpha_r y_r + \dots + \alpha_N y_N$$

splits into  $\Sigma_r$  and  $\bar{\Sigma}_{r-1}$  according to the condition  $\alpha_r = 0$  or  $\alpha_r > 0$ . It is therefore easy to ascertain a positive constant  $\epsilon_r \leq 1$  such that  $(1, y_{r+1}, \dots, y_N)$  lies in  $Z^{r-1}$  provided  $(y_{r+1}, \dots, y_N)$  lies in  $Z^r$  and

$$|y_{r+1}| \leq \epsilon_r, \dots, |y_N| \leq \epsilon_r.$$

This is true even at the first step  $\nu = 1$  when  $R^0 = R$  is restricted to  $G$ , because for a sufficiently small  $\epsilon$  the neighborhood of (75) described by

$$y_1 = 1, |y_2| \leq \epsilon, \dots, |y_N| \leq \epsilon$$

lies in  $G$ .

Starting with the point  $y_N = 0$  in  $Z^{N-1}$  and following this rule for the tran-

sition  $Z^r \rightarrow Z^{r-1}$  backwards from  $Z^{N-1}$  to  $Z$ , we find that the following  $N-1$  points

$$(1, 0, 0, \dots, 0),$$

$$(1, \epsilon_1, 0, \dots, 0),$$

$$(1, \epsilon_1, \epsilon_1 \epsilon_2, 0, \dots, 0),$$

.....

belong to  $Z$ . Thus the plane  $y_N = 0$  belongs to  $\Sigma_{N-1}$ , hence to  $\Sigma$ , is a face and contains the point  $f^1$ .

LEMMA 7. Any cell  $Z' = Z_S$  may be reached from the basic cell  $Z$  by a chain whose consecutive members are adjacent.

The inner reason for this lemma is obvious: because the region  $G$  is convex, the cell complex into which it has been divided is connected.

We start with the chain described in Lemma 5. Any two of its consecutive members have a common point  $f$  situated on the segment  $\tau$ ; but in general their contact will be one of order 1 only. We must insert further cells between them to make the chain proceed by contacts of order  $N-1$ .

The point  $f$ , being common to two cells, is not an inner point of a cell. I shall try to describe the situation intuitively in the plane section  $g_{nn} = 1$  of  $G$ . The cells to which  $f$  belongs cover an entire neighborhood  $U$  of  $f$ , each of them participating in it by an  $(N-1)$ -dimensional pyramid with vertex  $f$ . Hence we obtain a division of the  $(N-1)$ -dimensional space  $R^1$  into a finite number of convex pyramids radiating from the vertex  $f$ , and our task is to prove that this complex is connected. We thus face the same problem as before, but in one dimension less, and hence induction with respect to  $N$  will lead to the desired result. Let us now repeat the argument in detail, again using the notation  $y_1, \dots, y_N$  instead of  $g_{ij}$  for the coordinates in  $R$ .

Not more than a finite number of cells  $Z_S$  penetrate into a neighborhood  $U$  of  $f$  which lies in  $G(\rho, \sigma)$ . If one of these cells does not contain  $f$ , then  $U$  may be shrunk so as to have its intersection with the closed  $Z_S$  empty. Hence we find a smaller neighborhood of  $f$ , again called  $U$ , into which none but cells  $Z_f$  containing  $f$  will penetrate. We choose the coordinates  $y_i$  such that  $f = (1, 0, 0, \dots, 0)$ . A cell  $Z_f$  is defined by a finite set  $\Sigma$  of inequalities (74) which as before is divided into the subsets  $\Sigma_1$  and  $\bar{\Sigma}_0$ ; and as has been shown above, any point  $(1, y_2, \dots, y_N)$  sufficiently near to  $f$ , if it satisfies merely the inequalities  $\Sigma_1$ , will lie in  $Z_f$ . The inequalities  $\Sigma_1$  define a convex pyramid  $Z_f^{(1)}$  in the  $(N-1)$ -dimensional space  $R^1$  with the coordinates  $y_2, \dots, y_N$ . The center  $(y_1^0, \dots, y_N^0)$  of  $Z_f$  gives rise to a center  $(y_2^0, \dots, y_N^0)$  of  $Z_f^{(1)}$ . Thus the  $Z_f$  determine a division of  $R^1$  into a finite number of pyramids  $Z_f^{(1)}$ , and our aim is to prove the connectivity of that assemblage. Let us formulate this assertion as a lemma for  $N$  instead of  $N-1$  dimensions.

LEMMA 8<sub>N</sub>. Suppose the  $N$ -dimensional space  $R$  divided into a finite number of convex pyramids  $\Pi$  with their common vertex at the origin  $O$ . Each of them is supposed to contain inner points. Then any two of them can be joined by a chain whose consecutive members have contacts of order  $N-1$ .

The argument employed to reduce Lemma 7 to 8<sub>N-1</sub> may be used equally well to reduce 8<sub>N</sub> to 8<sub>N-1</sub> and thus to prove 8<sub>N</sub> by induction. The case is somewhat simpler because we now deal with a finite set of cells from the beginning. There is a slight complication, however, in so far as the Euclidean  $N$ -dimensional space robbed of the point  $O$  is not convex, but it is still connected as long as  $N \geq 2$ , and that is what counts. Indeed the centers of any two of our pyramids can be joined by a line consisting of one or two straight segments without passing through  $O$ .

As a consequence of Lemma 7, Theorem 14 is sharpened to

THEOREM 15. A complete system of modulo  $\{J\}$  incongruent substitutions  $S$  which carry  $Z$  into adjacent cells generates the whole modular group when one combines them with a system of generators for  $\{J\}$ .

13. Concluding remarks. Observe that a reduced form  $f$  satisfies the inequality

$$(77) \quad f(x_1, \dots, x_n) \geq g_{11}$$

not only for integers  $x_1, \dots, x_n$  without a (left) common divisor, but for any integers  $(x_1, \dots, x_n) \neq (0, \dots, 0)$  whatsoever. This is nothing else than the equation  $L_1 = M_1$ .

In this final section we are going to study the cell  $Z$  of reduced forms relatively to the whole  $N$ -dimensional space  $R$  rather than  $G$ .

The cell  $Z$  as a subset of  $R$  is not (necessarily) closed; boundary points  $f$  which do not belong to  $G$  will be semi-definite forms in the sense that  $f(\mathfrak{x}) \geq 0$  for every vector  $\mathfrak{x}$ , but  $f(\mathfrak{x}) = 0$  for certain vectors  $\mathfrak{x} \neq 0$ . Such a form can be written as a square sum

$$z_1^2 + \dots + z_m^2$$

of  $m < n$  linear forms  $z_1, \dots, z_m$  of the coordinates  $x_i$  with real coefficients. Now if  $\epsilon$  is any pre-assigned positive number we can ascertain a lattice vector  $(x_1, \dots, x_n) \neq (0, \dots, 0)$  for which

$$|z_1| \leq \epsilon, \dots, |z_m| \leq \epsilon$$

and thus

$$(78) \quad f \leq m\epsilon^2.$$

This is accomplished either by Minkowski's inequality (1) for a parallelo-  
tope or by an easy application of Dirichlet's principle concerning the distribu-

tion of  $\nu+1$  objects in  $\nu$  boxes. But (78) contradicts (77) unless

$$g_{11} = 0.$$

Because of the relations (63),

$$|g_{12}| \leq r g_{11}, \dots, |g_{1n}| \leq r g_{11},$$

which will extend from the forms in  $Z$  to those on the boundary of  $Z$ , the latter will satisfy the  $n$  equations

$$(79) \quad g_{11} = g_{12} = \dots = g_{1n} = 0.$$

(Even an appeal to the inequality  $g_{ii}^2 \leq g_{11} \cdot g_{ii}$  valid for all positive forms would have sufficed here.)

The closure  $\bar{Z}$  of  $Z$  in  $R$  has each of its boundary points either on one of the planes formerly assembled in the finite set  $\Sigma$  or on the plane  $g_{11}=0$ . Hence  $\bar{Z}$  as a part of  $R$  is completely described by the inequalities  $\Sigma$  together with  $g_{11} \geq 0$  and therefore is a pyramid. For  $n=1$ , the set  $\Sigma$  is empty and we have the one inequality  $g_{11} \geq 0$ . We may safely ignore this trivial case. For  $n \geq 2$ ,  $\bar{Z}$  reaches the boundary of  $G$  only along the "edge" (79) of  $n$  dimensions less; hence  $g_{11}=0$  is no face of  $\bar{Z}$ , and the inequality  $g_{11} \geq 0$  is redundant. Therefore:

**THEOREM 16.** *The same finite set of inequalities which defines  $Z$  in  $G$  defines  $\bar{Z}$  in  $R$ . The boundary points of  $\bar{Z}$  which do not belong to  $Z$  lie on the edge (79).*

The vertices of  $\bar{Z}$  are the so-called extreme forms; every reduced form is a linear combination of them with non-negative coefficients, but some of the extreme forms will be semi-definite.

We can now more fully appreciate the fine points in our two theorems of finiteness. By excluding from  $Z$  an arbitrarily small neighborhood

$$V_\epsilon: g_{11} < \epsilon g_{nn}$$

of the "edge" we obtain a compact subset  $Z_\epsilon$  of  $G^{(*)}$ . The fact that the boundary points of  $Z$  which lie outside this neighborhood  $V_\epsilon$  belong to a finite number of plane faces is considerably less deep than our first theorem of finiteness, and so is its proof. When one excludes  $V_\epsilon$ , one could have used the region  $G(\rho)$  defined by the first set of inequalities (61) alone instead of  $G(\rho, \sigma)$ , and could have shown that  $G(\rho)$  possesses not more than a finite number of plane faces outside  $V_\epsilon$ , while this is not true for  $G(\rho)$  or  $G(\rho, \sigma)$  as a whole. And the second theorem of finiteness could have been replaced by the less profound and more easily accessible assertion that there is only a finite number of  $S$  capable of carrying a point of  $Z$  outside of  $V_\epsilon$  into a point of  $G(\rho)$  outside  $V_\epsilon$ . These statements would have sufficed for the topological analysis

(\*) Compact under the convention that proportional forms like  $f$  and  $tf$  ( $t > 0$ ) are identified.



in §12. Our two theorems of finiteness include the approach to the "edge" and thus reveal finer features which are of great interest to the algebraist, though perhaps of less importance from the topological standpoint.

Up to now positive quadratic forms have been the object of investigation. Instead one can study arbitrary *affine coordinate systems* [16] in an  $n$ -dimensional vector space, consisting of  $n$  linearly independent vectors  $a_1, \dots, a_n$ ; these new objects form an  $n^2$ -dimensional space  $\mathfrak{A}$ . Two such systems  $(a_1, \dots, a_n)$  and  $(b_1, \dots, b_n)$  are said to be (arithmetically) equivalent if connected by a unimodular transformation  $S$ ,

$$b_i = \sum_k s_{ik} a_k \quad (s_{ik} \text{ integers, } \det(s_{ik}) = \pm 1).$$

For any vector  $\xi = (x_1, \dots, x_n)$  we introduce its square

$$\xi^2 = x_1^2 + \dots + x_n^2$$

(in accordance with Euclidean metric geometry) and associate the positive form

$$(80) \quad f(x_1, \dots, x_n) = (x_1 a_1 + \dots + x_n a_n)^2$$

with the coordinate system  $(a_1, \dots, a_n)^{(4)}$ . The latter is said to be reduced and to belong to the "cell"  $\mathfrak{J}$  of  $\mathfrak{A}$  provided the associated form  $f$  is reduced.  $\mathfrak{J}$  is a fundamental domain for the group  $\{S\}$  in  $\mathfrak{A}$ , and we could interpret our whole theory in terms of the new objects. The quadratic forms are then merely a tool for the study of coordinate systems under the rule of unimodular equivalence. We have thus returned to the approach of Chapter I: What we now call a reduced system  $(a_1, \dots, a_n)$  was there termed a reduced system with respect to the gauge function

$$(x_1^2 + \dots + x_n^2)^{1/2}.$$

A similar shift of viewpoint is applicable to the imaginary and the quaternion cases.

#### BIBLIOGRAPHY

1. Journal für die reine und angewandte Mathematik, vol. 129 (1905), pp. 220-274; also *Gesammelte Abhandlungen* II, Leipzig, 1911, pp. 53-100. Cited as M with the page number in the *Gesammelte Abhandlungen*.
2. Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1928, pp. 510-535; 1929, p. 508.
3. Quarterly Journal of Mathematics, vol. 9 (1938), pp. 259-262.
4. H. Weyl, *On geometry of numbers*, soon to appear in the Proceedings of the London Mathematical Society. On the whole subject see H. Hancock, *Development of the Minkowski Geometry of Numbers*, New York, 1939.

<sup>(4)</sup> The inequality (54),  $g_{11} \dots g_{nn} \geq D$ , for (80) reads in this interpretation as follows: The volume of a parallelotope cannot exceed the product of the lengths of the vectors by which it is spanned.



5. Another short proof by H. Davenport, *Quarterly Journal of Mathematics*, vol. 10 (1939), pp. 119-121.
6. *Compositio Mathematica*, vol. 5 (1938), pp. 368-391.
7. Cf. Minkowski's definition in M, p. 59.
8. See Mahler, loc. cit. (3 above), and the author, loc. cit. (4 above), Theorem V.
9. Weyl, loc. cit. (4 above), "Generalized Theorem V."
10. See M, pp. 56-58.
11. For more details see L. E. Dickson, *Algebren und ihre Zahlentheorie*, Zürich, 1927, chap. 9; C. G. Latimer, *American Journal of Mathematics*, vol. 48 (1926), pp. 57-66; M. Deuring, *Algebren, Ergebnisse der Mathematik*, vol. 4, no. 1, Berlin, 1935, chap. 6.
12. *Vorlesungen über die Zahlentheorie der Quaternionen*, Berlin, 1919.
13. The larger part of E. H. Moore's "Algebra of Matrices" (*General Analysis*, Part I, *Memoirs of the American Philosophical Society*, Philadelphia, 1935) deals with the formalism of "Hamiltonian" forms.
14. Cf. Weyl, loc. cit. (4 above), §8, and the more complicated argument in Bieberbach-Schur, loc. cit. (2 above), pp. 521-523.
15. Loc. cit. (6 above), equation (25).
16. See M, p. 53.

INSTITUTE FOR ADVANCED STUDY,  
PRINCETON, N. J.

# CONTINUED FRACTIONS AND TOTALLY MONOTONE SEQUENCES

BY  
H. S. WALL

1. **Introduction.** A sequence  $c_0, c_1, c_2, \dots$  of real numbers is called *totally monotone* if  $\Delta^m c_n \geq 0$ , ( $m, n = 0, 1, 2, \dots$ ), where

$$\Delta^m c_n = c_n - C_{m,1}c_{n+1} + C_{m,2}c_{n+2} - \dots + (-1)^m C_{m,m}c_{n+m}.$$

Hausdorff<sup>(1)</sup> showed that for every totally monotone sequence  $c_0, c_1, c_2, \dots$  there exists (essentially uniquely) a monotone nondecreasing real function  $\phi(u)$ ,  $0 \leq u \leq 1$ , such that

$$c_n = \int_0^1 u^n d\phi(u), \quad n = 0, 1, 2, \dots.$$

Conversely, if  $\phi(u)$  is a monotone nondecreasing bounded real function on the interval  $0 \leq u \leq 1$ , then  $\Delta^m c_n = \int_0^1 (1-u)^m u^n d\phi(u) \geq 0$ ,  $m, n = 0, 1, 2, \dots$ , so that  $c_0, c_1, c_2, \dots$  is totally monotone.

In case the function  $\phi(u)$  has an infinity of points of increase in the interval  $0 \leq u \leq 1$ , then the corresponding sequence is a special *Stieltjes moment sequence*, and accordingly there is a Stieltjes continued fraction<sup>(2)</sup>

$$(1.1) \quad b_1/1 + b_2x/1 + b_3x/1 + \dots,$$

in which the numbers  $b_1, b_2, b_3, \dots$  are real and positive, which corresponds to the power series

$$(1.2) \quad c_0 - c_1x + c_2x^2 - \dots.$$

On the other hand, if  $\phi(u)$  has but a finite number of points of increase, then the series (1.2) represents a rational function of  $x$  and the continued fraction terminates.

The main problem which we have solved in the present paper is as follows: *to find necessary and sufficient conditions upon the numbers  $b_1, b_2, b_3, \dots$  in the continued fraction (1.1) in order that the coefficients  $c_0, c_1, c_2, \dots$  in the corresponding power series (1.2) shall form a totally monotone sequence.* The re-

Presented to the Society February 24, 1940; received January 30, 1940. This paper is dedicated to Edward Burr Van Vleck on the occasion of his seventy-seventh birthday, June 7, 1940.

(<sup>1</sup>) F. Hausdorff, *Ueber das Momentenproblem für ein endliches Intervall*, Mathematische Zeitschrift, vol. 16 (1923), pp. 220-248.

(<sup>2</sup>) T. J. Stieltjes, *Recherches sur les fractions continues*, Oeuvres, vol. 2, pp. 402-566. We have made the substitution of  $x$  for  $1/x$ , and have put  $b_1 = 1/a_1$ ,  $b_n = 1/a_n a_{n-1}$ ,  $n \geq 2$ , in the series and continued fraction used by Stieltjes.

sult is very simple, namely: the sequence  $c_0, c_1, c_2, \dots$  is totally monotone if and only if there exist real numbers  $g_0, g_1, g_2, \dots$  such that  $0 \leq g_n \leq 1$ ,  $n=0, 1, 2, \dots$ , and such that

$$c_0 - c_1x + c_2x^2 - c_3x^3 + \dots \sim g_0/1 + g_1x/1 + (1 - g_1)g_2x/1 \\ + (1 - g_2)g_3x/1 + \dots,$$

it being agreed that the continued fraction<sup>(3)</sup> shall terminate with the first identically vanishing partial quotient.

If  $c_0, c_1, c_2, \dots$  is totally monotone, the function  $f(x)$  represented by the power series (1.2) is analytic for  $|x| < 1$ . Let  $M(f) = \text{l.u.b.}_{|x| < 1} |f(x)|$ . We show that  $M(f)$  is finite if and only if the series  $c_0 + c_1 + c_2 + \dots$  converges, and establish the equality

$$M(f) = c_0 + c_1 + c_2 + \dots.$$

We also characterize the class  $E$  of these "moment generating functions" which are bounded in the unit circle (a) in terms of the Stieltjes integral representation of  $f(x)$ , and (b) in terms of the continued fraction representation of  $f(x)$ . It is shown that if  $f(x) \in E$ ,  $M(f) \leq 1$ , then the functions defined by the algorithm of Schur<sup>(4)</sup>, namely:

$$f_{n+1} = \frac{1}{x} \frac{t_n - f_n}{1 - t_n f_n}, \quad t_n = f_n(0),$$

$n=0, 1, 2, \dots$ ,  $f_0 = f$ , are all in  $E$  and have moduli not exceeding 1 for  $|x| < 1$ .

2. **An operation on continued fractions.** We shall use the symbol " $\sim$ " between a power series  $P(x)$  and a continued fraction  $K(x)$  to indicate that the power series expansion of the  $n$ th approximant of  $K(x)$  agrees term by term with  $P(x)$  for more and more terms as  $n$  is increased, or becomes identical with  $P(x)$  from and after some value of  $n$ . The basic theorem of the paper is

**THEOREM 2.1.** *If  $g_1, g_2, g_3, \dots$  are any real or complex numbers, and  $P(x)$  is a power series in ascending powers of  $x$  such that*

$$(2.1) \quad P(x) \sim 1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots,$$

*then*

$$(2.2) \quad \frac{1+x}{P(x)} \sim 1 + (1 - g_1)x/1 + g_1(1 - g_2)x/1 + g_2(1 - g_3)x/1 + \dots.$$

<sup>(3)</sup> Continued fractions of this form were first treated by E. B. Van Vleck, in a paper entitled *On the convergence and character of the continued fraction  $a_1x/1 + a_2x/1 + a_3x/1 + \dots$* , these Transactions, vol. 2 (1901), pp. 476-483.

<sup>(4)</sup> J. Schur, *Ueber Potenzreihen, die im Innern des Einheitskreises beschränkt sind*, Journal für die reine und angewandte Mathematik, vol. 147 (1916), pp. 205-232, and vol. 148 (1917), pp. 122-145.

**Proof.** Let  $A_n(x)/B_n(x)$ ,  $A_n^*(x)/B_n^*(x)$  be the  $n$ th approximants of the continued fractions in (2.1) and (2.2), respectively. Then we have the relations

$$(2.3) \quad A_n(x) = g_n x B_{n-1}^*(x) + B_n^*(x), \quad (1+x)B_n(x) = g_n x A_{n-1}^*(x) + A_n^*(x),$$

$(n=0, 1, 2, \dots, g_0=1, A_{-1}^*=1, B_{-1}^*=0)$ . These may be verified directly for  $n=0, 1$ . Assuming that the first is true for  $n \leq m$ ,  $m \geq 1$ , we then have

$$\begin{aligned} A_{m+1}(x) &= A_m(x) + g_{m+1}(1-g_m)x A_{m-1}(x) \\ &= g_m x B_{m-1}^*(x) + B_m^*(x) + g_{m+1}(1-g_m)x [g_{m-1}x B_{m-2}^*(x) + B_{m-1}^*(x)]. \end{aligned}$$

Since  $g_{m-1}(1-g_m)x B_{m-2}^*(x) = B_m^*(x) - B_{m-1}^*(x)$ , we then have

$$\begin{aligned} A_{m+1}(x) &= g_{m+1}x B_m^*(x) + [B_m^*(x) + g_m(1-g_{m+1})x B_{m-1}^*(x)] \\ &= g_{m+1}x B_m^*(x) + B_{m+1}^*(x), \end{aligned}$$

so that the first relation (2.3) is true for  $n=m+1$ , and therefore, by mathematical induction, for all  $n$ . The second relation (2.3) may be proved in a similar way.

On multiplying the first relation (2.3) by  $A_{n-1}^*(x)$ , the second by  $B_{n-1}^*(x)$ , and then subtracting, we find that the power series expansion in ascending powers of  $x$  of the rational function

$$A_{n-1}^*(x)/B_{n-1}^*(x) - \frac{1+x}{A_n(x)/B_n(x)}$$

begins with the  $(n-1)$ th or a higher power of  $x$ . It follows immediately that the correspondence (2.1) implies the correspondence (2.2).

This theorem may be thrown into the following form:

**THEOREM 2.2.** Let  $c_0 \neq 0$ , and

$$(2.4) \quad c_0 - c_1x + c_2x^2 - \dots \sim c_0/1 + g_1x/1 + (1-g_1)g_2x/1 + (1-g_2)g_3x/1 + \dots$$

Then

$$(2.5) \quad \Delta c_0 - \Delta c_1x + \Delta c_2x^2 - \dots \sim \Delta c_0/1 + g_1(1-g_2)x/1 + g_2(1-g_3)x/1 + \dots$$

**Proof.** Let  $c_0/P(x) = c_0 - c_1x + c_2x^2 - \dots$  in Theorem 2.1. This gives at once

$$\begin{aligned} c_0 + (c_0 - c_1)x - (c_1 - c_2)x^2 + \dots &\sim c_0 + c_0(1-g_1)x/1 \\ &\quad + g_1(1-g_2)x/1 + \dots \end{aligned}$$

On removing the constant term  $c_0$  from the series and from the continued fraction, and then dropping a factor  $x$ , the correspondence (2.5) results. We note for future reference that

$$(2.6) \quad \Delta c_0 = c_0(1 - g_1).$$

By means of these theorems we have enlarged by one the small list of known operations on continued fractions.

The next theorem makes the transformation available for a large class of continued fractions.

**THEOREM 2.3.** *Let  $A_n(x)/B_n(x)$  be the  $n$ th approximant of the continued fraction*

$$(2.7) \quad 1 + a_1x/1 + a_2x/1 + a_3x/1 + \dots,$$

*and let  $c \neq 0$  be any number such that  $A_n(-c) \neq 0$ , ( $n = 1, 2, 3, \dots$ ). Put*

$$(2.8) \quad g_n = a_n c A_{n-2}(-c) / A_{n-1}(-c),$$

*$n = 1, 2, 3, \dots$ ,  $A_{-1}(-c) = 1$ . Then the continued fraction (2.7) takes the form*

$$1 + \frac{g_1(x/c)}{1} + \frac{(1 - g_1)g_2(x/c)}{1} + \frac{(1 - g_2)g_3(x/c)}{1} + \dots$$

**Proof.** One may verify immediately that  $a_1 = g_1/c$ ,  $a_n = g_n(1 - g_{n-1})/c$ ,  $n = 2, 3, 4, \dots$ , when the  $g_n$ 's are given by (2.8).

We remark in passing that if  $a_n \neq 0$  in (2.7), and  $P(x)$  is the power series corresponding to (2.7), then one may apply Theorem 2.1 to obtain the formulas given by Stieltjes<sup>(6)</sup> for the continued fraction corresponding to  $1/P(x)$ . In fact:

$$(2.9) \quad \frac{1 + (x/c)}{P(x)} \sim 1 + \frac{(1 - g_1)(x/c)}{1} + \frac{g_2(1 - g_1)(x/c)}{1} + \frac{g_3(1 - g_2)(x/c)}{1} + \dots$$

We may allow  $c$  to become infinite and the correspondence will be maintained provided the coefficients of  $x$  in the continued fraction have limits which are finite and not 0. Accordingly we obtain this theorem:

**THEOREM 2.4.** *If the power series  $P(x)$  has a corresponding continued fraction (2.7) in which  $a_n \neq 0$ ,  $n \geq 1$ , then  $1/P(x)$  will have a corresponding continued fraction of the same form provided  $A_{2n}(x)$ , the numerator of the  $2n$ th approximant of (2.7), is of degree  $n$  for  $n = 0, 1, 2, \dots$ .*

The condition of the theorem is met when the  $a_n$ 's are real and positive, which is the case with which Stieltjes was concerned. If one evaluates the limits for  $c = \infty$  of the coefficients of  $x$  in the continued fraction of (2.9), he will find that the result agrees with that found by Stieltjes.

<sup>(6)</sup> O. Perron, *Die Lehre von den Kettenbrüchen*, 1st edition, pp. 334-335.

3. **A uniform convergence theorem.** In a recent<sup>(6)</sup> paper Scott and Wall proved a theorem which may be stated in the following form:

**THEOREM 3.1.** *If  $g_1, g_2, g_3, \dots$  are real numbers such that  $0 < g_1 < 1$ ,  $0 \leq g_n < 1$ ,  $n > 1$ , and  $x_1, x_2, x_3, \dots$  are functions of any variables, then the continued fraction*

$$(3.1) \quad \frac{1}{1 + \frac{(1 - g_1)g_2x_1}{1 + \frac{(1 - g_2)g_3x_2}{1 + \frac{(1 - g_3)g_4x_3}{1 + \dots}}}}$$

*converges uniformly for  $|x_n| \leq 1$ ,  $n = 1, 2, 3, \dots$ . The denominators of all the approximants are nonzero in this domain. Let  $G$  denote the value of the continued fraction. Then*

$$|G| \leq \frac{1}{g_1} \left\{ 1 - \frac{1}{1 + \sum_{n=1}^{\infty} \frac{g_1 g_2 \cdots g_n}{(1 - g_1)(1 - g_2) \cdots (1 - g_n)}} \right\}$$

*if  $|x_n| \leq 1$ ,  $n \geq 1$ ; and  $G$  is equal to the expression on the right if  $x_n = -1$ ,  $n \geq 1$ .*

We shall digress momentarily at this point to discuss two theorems given by Perron on page 262 of his book. The first of these may, with no essential loss in generality, be stated as follows:

**THEOREM 3.2.** *If the elements of the continued fraction*

$$(3.2) \quad 1/1 + a_2/1 + a_3/1 + a_4/1 + \dots$$

*are functions of any variables, then the continued fraction converges uniformly over the domain characterized by the inequalities*

$$|a_n| \leq (p_n - 1)/p_n p_{n-1}, \quad n = 2, 3, 4, \dots,$$

*where  $p_1, p_2, p_3, \dots$  are any real constants greater than 1 for which the series*

$$(3.3) \quad \sum_{n=1}^{\infty} (p_1 - 1)(p_2 - 1) \cdots (p_n - 1)$$

*is divergent.*

Perron then says: "Ein bemerkenswerter Spezialfall unseres allgemeinen Kriteriums ist" and then *proves* a theorem, attributed to Van Vleck, which may be stated as follows:

**THEOREM 3.3.** *If  $g_1, g_2, g_3, \dots$  are real numbers such that  $0 < g_n < 1$ ,  $n \geq 1$ , and  $x_1, x_2, x_3, \dots$  are functions of any variables, then the continued fraction (3.1) converges uniformly for  $|x_n| \leq 1$ ,  $n \geq 1$ .*

<sup>(6)</sup> W. T. Scott and H. S. Wall, *A convergence theorem for continued fractions*, these Transactions, vol. 47 (1940), pp. 155-172.



It is a strange fact that *the second theorem is more general than the first*. To see this, put  $a_n = g_n(1 - g_{n-1})x_n$ ,  $n \geq 2$ ;  $p_n = 1/(1 - g_n)$ ,  $n \geq 1$ , and the second theorem reduces to the first *minus* the requirement on the series (3.3).

It should be added that Theorem 3.3 is related to but quite different from the theorem which Van Vleck proved<sup>(7)</sup>. He gave preference to the continued fraction  $1/1 + g_1x_0/1 + (1 - g_1)g_2x_1/1 + (1 - g_2)g_3x_2/1 + \dots$ , the reciprocal of which is, except for an unimportant term and factor, the continued fraction (3.1). Theorem 3.1 is an improvement over Theorem 3.3, in that the  $g_n$ 's after the first are permitted to be 0. For this reason Theorem 3.1 contains the theorem given by Perron on page 258 (Theorem 26), in which it may be assumed with no loss in generality that  $p_1 > 1$ .

**4. Totally monotone sequences corresponding to an "infinite distribution of mass."** The sequence  $c_n = \int_0^1 u^n d\phi(u)$ ,  $n = 0, 1, 2, \dots$ , in which  $\phi(u)$  is real and monotone nondecreasing is completely characterized by the inequalities

$$(4.1) \quad \Delta^m c_n \geq 0, \quad m, n = 0, 1, 2, \dots,$$

and is said to be a *totally monotone* sequence. If  $\phi(u)$  has an infinite number of points of increase, we shall say that there is an *infinite distribution of mass*.

**THEOREM 4.1.** *The sequence  $c_0, c_1, c_2, \dots$  is a totally monotone sequence corresponding to an infinite distribution of mass if and only if there exist real numbers  $g_1, g_2, g_3, \dots$  such that  $0 < g_n < 1$ ,  $n \geq 1$ , and such that*

$$(4.2) \quad c_0 - c_1x + c_2x^2 - \dots \sim c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots, \quad (c_0 > 0).$$

**Proof.** Suppose that (4.2) holds. Then by Theorem 3.1 the continued fraction  $1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$  converges uniformly for  $|x| \leq 1$ . If  $f(x)$  is the analytic function represented, then by Theorem 2.1 and Theorem 3.1,  $(1+x)/f(x)$  is analytic for  $|x| < 1$ , and therefore  $c_0/f(x)$ , the function represented by the continued fraction and series (4.2), is analytic for  $|x| < 1$ .

Now the coefficients of  $x$  in the continued fraction (4.2) are positive, and hence by the work of Stieltjes, this continued fraction represents a function of the form  $\int_0^1 d\phi(u)/(1+xu)$ , where  $\phi(u)$  is monotone nondecreasing, and has an infinite number of points of increase. Inasmuch as this function is analytic for  $|x| < 1$ , and the corresponding continued fraction converges uniformly for  $|x| \leq r$  where  $r$  is any positive number less than 1, it follows that the upper limit of integration may be taken equal to 1. Then  $\Delta^m c_n = \int_0^1 (1-u)^m u^n d\phi(u)$ , and therefore (4.1) holds. Thus the sequence is totally monotone, and corresponds to an infinite distribution of mass.

Conversely, let  $c_n = \int_0^1 u^n d\phi(u)$ , where  $\phi(u)$  is monotone nondecreasing and has an infinite number of points of increase. That is,  $c_0, c_1, c_2, \dots$  is a totally

<sup>(7)</sup> See footnote 3.

monotone sequence corresponding to an infinite distribution of mass. Then, by the work of Stieltjes, we must have a correspondence of the form

$$c_0 - c_1x + c_2x^2 - \dots \sim a_1/1 + a_2x/1 + a_3x/1 + \dots,$$

where  $a_1, a_2, a_3, \dots$  are real and positive. Moreover, the function represented by this series and continued fraction is  $\int_0^1 d\phi(u)/(1+xu)$ . Since the limits of integration are from 0 to 1, the zeros<sup>(8)</sup> of  $B_n(x)$ , the denominator of the  $n$ th approximant of the continued fraction, are real and less than  $-1$ . Since  $B_n(0)=1$  it therefore follows that  $B_n(-1)>0$ . We may then apply Theorem 2.3, with  $c=1$ , to the continued fraction

$$1 + a_2x/1 + a_3x/1 + a_4x/1 + \dots,$$

the numerator of whose  $n$ th approximant is  $B_{n+1}(x)$ , and thus determine numbers  $g_1, g_2, g_3, \dots$ , such that

$$a_2 = g_1, \quad a_n = g_{n-1}(1 - g_{n-2}), \quad n > 2.$$

Now, by Theorem 2.2,

$$\Delta c_0 - \Delta c_1x + \Delta c_2x^2 - \dots \sim \frac{\Delta c_0}{1} + \frac{g_1(1 - g_2)x}{1} + \frac{g_2(1 - g_3)x}{1} + \dots.$$

Also,  $\Delta c_n = \int_0^1 u^n d\phi_1(u)$  where  $\phi_1(u) = \int_0^u (1-u)d\phi(u)$  is monotone nondecreasing and has an infinite number of points of increase. It follows that the coefficients of  $x$  in the last continued fraction must all be positive, and that  $\Delta c_0 = c_0(1 - g_1) > 0$  (cf. (2.6)). We therefore have

$$\begin{aligned} g_1 &> 0, & g_{n-1}(1 - g_{n-2}) &> 0, & n > 2, \\ 1 - g_1 &> 0, & g_{n-2}(1 - g_{n-1}) &> 0, & n > 2, \end{aligned}$$

and consequently  $0 < g_n < 1$ ,  $n=1, 2, 3, \dots$ , as was to be proved.

**5. Developments from the continued fraction algorithm.** We shall begin by considering an example. It is known<sup>(9)</sup> that the series

$$F(\alpha, 1, \gamma; -x) = 1 - \frac{\alpha}{\gamma}x + \frac{\alpha(\alpha+1)}{\gamma(\gamma+1)}x^2 - \dots$$

has the corresponding continued fraction  $1 + e_1x/1 + e_2x/1 + e_3x/1 + \dots$ , where

$$e_1 = \frac{\alpha}{\gamma}, \quad e_{2n} = \frac{(\alpha+n)(\gamma+n-1)}{(\gamma+2n-2)(\gamma+2n-1)}, \quad e_{2n+1} = \frac{n(\gamma+\alpha+n-1)}{(\gamma+2n-1)(\gamma+2n)}.$$

Naturally  $\alpha, \gamma$  are not negative integers or 0. We then readily find that

<sup>(8)</sup> Perron, loc. cit., p. 368 and p. 383.

<sup>(9)</sup> Perron, loc. cit., p. 348.

$$F(\alpha, 1, \gamma; -x) = \frac{1}{1+x} + \frac{g_1 x}{1} + \frac{(1-g_1)g_2 x}{1} + \frac{(1-g_2)g_3 x}{1} + \dots,$$

where  $g_{2n} = n/(\gamma + 2n - 1)$ ,  $g_{2n-1} = (\alpha + n - 1)/(\gamma + 2n - 2)$ ,  $n = 1, 2, 3, \dots$ . On applying Theorem 2.1 we readily obtain the power series identity

$$(5.1) \quad F(\alpha, 1, \gamma; -x) = \frac{1}{1+x} + \frac{\gamma - \alpha}{\gamma} \frac{x}{1+x} F(\alpha, 1, \gamma + 1; -x).$$

The repeated application of this identity gives the Euler expansion

$$\begin{aligned} F(\alpha, 1, \gamma; -x) &= \frac{1}{1+x} + \frac{(\gamma - \alpha)}{\gamma} \frac{x}{(1+x)^2} \\ &\quad + \frac{(\gamma - \alpha)(\gamma + 1 - \alpha)}{\gamma(\gamma + 1)} \frac{x^2}{(1+x)^3} + \dots \end{aligned}$$

We now propose to obtain the analogous developments for the general continued fraction of this form. For simplicity, let  $g_n$  be real and  $0 < g_n < 1$ , and put

$$f(x) = 1 + g_1 x/1 + (1 - g_1)g_2 x/1 + (1 - g_2)g_3 x/1 + \dots,$$

so that

$$(1+x)/f(x) = 1 + (1 - g_1)x/1 + g_1(1 - g_2)x/1 + g_2(1 - g_3)x/1 + \dots$$

Let  $f_1(x) = 1 + g_1(1 - g_2)x/1 + g_2(1 - g_3)x/1 + \dots$ , and denote by  $A_n(x)$  the numerator of the  $n$ th approximant of this continued fraction. Then it is easy to prove by mathematical induction that

$$A_{n-1}(-1) = g_1 g_2 \dots g_n S_n,$$

where

$$S_n = 1 + \sum_{k=1}^n \frac{(1 - g_1)(1 - g_2) \dots (1 - g_k)}{g_1 g_2 \dots g_k}.$$

We may therefore apply Theorem 2.3 with  $c=1$  and obtain

$$f_1(x) = 1 + g_1^{(1)} x/1 + g_2^{(1)} (1 - g_1^{(1)}) x/1 + g_3^{(1)} (1 - g_2^{(1)}) x/1 + \dots,$$

where  $g_n^{(1)} = g_n(1 - g_{n+1})A_{n-2}(-1)/A_{n-1}(-1) = (1 - g_{n+1})S_{n-1}/S_n$ ,  $n = 1, 2, 3, \dots$ , ( $S_0 = 1$ ). Consequently  $0 < g_n^{(1)} < 1$ , so that the continued fraction for  $f_1(x)$  has precisely the same form as that for  $f(x)$ . Hence we may write

$$\frac{1+x}{f_1(x)} = 1 + \frac{(1 - g_1^{(1)})x}{1} + \frac{g_1^{(1)}(1 - g_2^{(1)})x}{1} + \frac{g_2^{(1)}(1 - g_3^{(1)})x}{1} + \dots,$$

put  $f_2(x) = 1 + g_1^{(1)}(1 - g_2^{(1)})x/1 + g_2^{(1)}(1 - g_3^{(1)})x/1 + \dots$ , and apply the above argument to the latter continued fraction. In this manner we arrive at the following theorem:

**THEOREM 5.1.** Let  $f(x) = 1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$ , where  $0 < g_n < 1$ . Define numbers  $g_n^{(m)}$  by the relations

$$(5.2) \quad g_k^{(0)} = g_k, \quad g_k^{(m)} = (1 - g_{k+1}^{(m-1)})S_{k-1}^{(m-1)}/S_k^{(m-1)},$$

$m, k = 1, 2, 3, \dots$ ,

$$(5.3) \quad S_0^{(k)} = 1, \quad S_i^{(k)} = 1 + \sum_{j=1}^i \frac{(1 - g_1^{(k-1)})(1 - g_2^{(k-1)}) \dots (1 - g_j^{(k-1)})}{g_1^{(k-1)}g_2^{(k-1)} \dots g_j^{(k-1)}},$$

so that  $0 < g_n^{(m)} < 1$ ; and put  $f_k(x) = 1 + g_1^{(k)}x/1 + (1 - g_1^{(k)})g_2^{(k)}x/1 + (1 - g_2^{(k)})g_3^{(k)}x/1 + \dots$ . Then the functions  $f_0 \equiv f, f_1, f_2, f_3, \dots$  satisfy the following identities:

$$\frac{1}{f_k} = \frac{1}{1+x} + (1 - g_1^{(k)}) \frac{x}{1+x} \frac{1}{f_{k+1}},$$

$k = 0, 1, 2, \dots$ , and consequently  $1/f(x)$  has the Euler expansion

$$\frac{1}{f(x)} = \frac{1}{1+x} + (1 - g_1) \frac{x}{(1+x)^2} + (1 - g_1)(1 - g_1^{(1)}) \frac{x^2}{(1+x)^3} + \dots$$

**COROLLARY 5.1.** Let  $c_0 \neq 0$ , and  $c_0/f(x) = c_0 - c_1x + c_2x^2 - \dots$ . Then

$$(5.4) \quad \Delta^m c_0 = c_0(1 - g_1)(1 - g_1^{(1)}) \dots (1 - g_1^{(m-1)}).$$

**COROLLARY 5.2.** The correspondence

$$(5.5) \quad \Delta^m c_0 - \Delta^m c_1 x + \Delta^m c_2 x^2 - \dots \sim \frac{\Delta^m c_0}{1} + \frac{g_1^{(m)} x}{1} + \frac{(1 - g_1^{(m)})g_2^{(m)} x}{1} + \frac{(1 - g_2^{(m)})g_3^{(m)} x}{1} + \dots$$

is valid for  $m = 0, 1, 2, \dots$ .

From (5.4) it follows that if  $c_0 > 0$  then  $\Delta^m c_0 > 0$  for  $m = 1, 2, 3, \dots$ . We shall show next that  $\Delta^m c_n > 0$  for all  $m, n$ . To do this it will suffice to prove

**THEOREM 5.2.** If  $0 < g_n < 1$  and

$$c_0 - c_1x + c_2x^2 - \dots \sim c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots,$$

then

$$c_1 - c_2x + c_3x^2 - \dots \sim c_1/1 + h_1x/1 + (1 - h_1)h_2x/1$$

$$+ (1 - h_2)h_3x/1 + \dots,$$

where  $0 < h_n < 1$ .

**Proof.** With the aid of Stieltjes integrals and the results of §4, one may prove this theorem in a few lines. The proof may be made by means of the continued fraction algorithm alone by means of the four lemmas which follow.

LEMMA 5.1. *The continued fraction*<sup>(10)</sup>

$$u_1 + v_1 - \frac{u_1v_2}{u_2 + v_2} - \frac{u_2v_3}{u_3 + v_3} - \dots - \frac{u_{n-1}v_n}{u_n + v_n}$$

in which  $u_i \neq 0$ ,  $i = 1, 2, 3, \dots, n$ , is equal to

$$v_1 + \frac{u_1}{1 + \frac{v_2}{u_2} + \frac{v_2v_3}{u_2u_3} + \dots + \frac{v_2v_3 \dots v_n}{u_2u_3 \dots u_n}}$$

This may be readily proved by mathematical induction.

LEMMA 5.2. *If  $a_n > 0$  and*

$$c_0 - c_1x + c_2x^2 - \dots \sim a_1/1 + a_2x/1 + a_3x/1 + \dots,$$

then<sup>(11)</sup>

$$c_1 - c_2x + c_3x^2 - \dots \sim b_1/1 + b_2x/1 + b_3x/1 + \dots,$$

where

$$(5.6) \quad b_1 = a_1a_2, \quad b_2 = a_2 + a_3, \quad b_{2n+1} = a_{2n+3} + a_{2n+2} - b_{2n+2},$$

$$(5.7) \quad b_{2n+2} = a_{2n+3} + \frac{a_{2n+2}}{1 + \frac{a_{2n+1}}{a_{2n}} + \frac{a_{2n+1}a_{2n-1}}{a_{2n}a_{2n-2}} + \dots + \frac{a_{2n+1}a_{2n-1} \dots a_3}{a_{2n}a_{2n-2} \dots a_2}}$$

**Proof.** It is well known that  $c_1 - c_2x + c_3x^2 - \dots$  has a corresponding continued fraction of the form specified. Moreover, the "odd part" of  $a_1/1 + a_2x/1 + a_3x/1 + \dots$  must be the same as the "even part" of  $b_1/1 + b_2x/1 + b_3x/1 + \dots$ . That is,

$$(5.8) \quad \begin{aligned} a_1 - \frac{a_1a_2x}{1 + (a_2 + a_3)x} - \frac{a_3a_4x^2}{1 + (a_4 + a_5)x} - \frac{a_5a_6x^2}{1 + (a_6 + a_7)x} - \dots \\ = c_0 - \frac{b_1x}{1 + b_2x} - \frac{b_2b_3x^2}{1 + (b_3 + b_4)x} - \frac{b_4b_5x^2}{1 + (b_5 + b_6)x} - \dots \end{aligned}$$

<sup>(10)</sup> O. Szász, *Ueber die Erhaltung der Konvergenz unendlicher Kettenbrüche* ..., Journal für die reine und angewandte Mathematik, vol. 147 (1916), pp. 132-160.

<sup>(11)</sup> This holds for any  $a_n$ 's such that there is no division by zero in the formulas.

On equating corresponding elements in these continued fractions we obtain (5.6) and the relation  $b_{2n}b_{2n+1} = a_{2n+1}a_{2n+2}$ . Hence, on combining this with (5.6),

$$b_{2n+2} = a_{2n+2} + a_{2n+2} - \frac{a_{2n+1}a_{2n+2}}{a_{2n+1} + a_{2n}} - \frac{a_{2n-1}a_{2n}}{a_{2n-1} + a_{2n-2}} - \cdots - \frac{a_3a_4}{a_2 + a_3},$$

which, by Lemma 5.1, reduces to (5.7).

**LEMMA 5.3.** *If in Lemma 5.2 the  $a_n$ 's have the form  $a_2 = g_1$ ,  $a_n = g_{n-1}(1 - g_{n-2})$ ,  $n > 2$ , where  $0 < g_n < 1$ , then  $0 < b_n < 1$ ,  $n \geq 2$ .*

**Proof.** We have  $0 < b_2 = 1 - (1 - g_1)(1 - g_2) < 1$ ,  $0 < b_{2n+2} = 1 - (1 - g_{2n+1})(1 - g_{2n+2}) - g_{2n+1}g_{2n} < 1$ . Hence also,  $0 < b_{2n+1} < 1$ .

**LEMMA 5.4.** *Under the hypothesis of Lemma 5.3 the numerators of the approximants of the continued fraction  $1 + b_2x/1 + b_3x/1 + b_4x/1 + \cdots$  are all positive for  $x = -1$ .*

**Proof.** Let  $A_n(x)$ ,  $n = 0, 1, 2, \dots$ , be the numerators in question, and let  $C_n(x)$  be the denominator of the  $n$ th approximant of

$$(5.9) \quad a_1/1 + a_2x/1 + a_3x/1 + \cdots.$$

Then by (5.8),  $C_{2n+1}(x) \equiv A_{2n}(x)$ . Now  $|C_2(x)| = |1 + g_1x| \geq (1 - g_1)|C_1(x)|$ , if  $|x| \leq 1$ . If, for any value of  $n$ ,  $|x| \leq 1$  implies that

$$|C_n(x)| \geq (1 - g_{n-1})|C_{n-1}(x)| \geq (1 - g_1)(1 - g_2) \cdots (1 - g_{n-1}),$$

then  $|C_{n+1}(x)| = |C_n(x) + g_n(1 - g_{n-1})x C_{n-1}(x)| \geq |C_n(x)| - g_n(1 - g_{n-1})|C_{n-1}(x)|$  if  $|x| \leq 1$ . Hence<sup>(12)</sup>

$$|C_{n+1}(x)| \geq (1 - g_n)|C_n(x)| \geq (1 - g_1)(1 - g_2) \cdots (1 - g_n).$$

Inasmuch as  $C_n(0) = 1$ , we must therefore have  $C_{2n+1}(-1) = A_{2n}(-1) > 0$ . Now  $A_{2n+2}(x) = A_{2n+1}(x) + b_{2n+2}x A_{2n}(x)$ . Hence if  $A_{2n+1}(x) = 0$  for  $-1 \leq x \leq 0$ , then  $A_{2n+2}(x)$  and  $A_{2n}(x)$  would have opposite signs for this value of  $x$ . Since this is impossible, it follows that  $A_{2n+1}(-1) > 0$ .

The proof of Theorem 5.2 may now be readily made. By Lemma 5.4 we may apply Theorem 2.3 with  $c = 1$  and obtain  $b_2 = h_1$ ,  $b_n = h_{n-1}(1 - h_{n-2})$ ,  $n > 2$ , where  $h_n = b_{n+1}A_{n-2}(-1)/A_{n-1}(-1) > 0$ . But by Lemma 5.3,  $0 < h_1 < 1$ ,  $0 < h_{n-1}(1 - h_{n-2}) < 1$ ,  $m > 2$ . Consequently  $0 < h_n < 1$ ,  $n \geq 1$ .

**THEOREM 5.3.** *If  $c_0 \geq 0$ , and*

$$c_0 - c_1x + c_2x^2 - \cdots \sim c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \cdots,$$

where  $0 < g_n < 1$ , then

<sup>(12)</sup> This induction was used by Van Vleck, loc. cit.



$$(5.10) \quad \Delta^m c_n > 0, \quad m, n = 0, 1, 2, \dots,$$

and

$$(5.11) \quad a < \frac{\Delta c_n}{c_n} < \frac{\Delta^2 c_n}{\Delta c_n} < \frac{\Delta^3 c_n}{\Delta^2 c_n} < \dots < 1, \quad n = 0, 1, 2, \dots$$

**Proof.** The inequalities (5.10) follow from (5.4) and Theorem 5.2. It will suffice to prove (5.11) for the case  $n=0$ . By (5.4) we have

$$\begin{aligned} 1 &> \frac{\Delta^{m+1} c_0}{\Delta^m c_0} = (1 - g_1^{(m)}) = 1 - g_1^{(m-1)} (1 - g_2^{(m-1)}) \\ &> 1 - g_1^{(m-1)} - \frac{\Delta^m c_0}{\Delta^{m-1} c_0} > 0, \end{aligned}$$

$m=1, 2, 3, \dots$ , which was to be proved.

**6. Totally monotone sequences corresponding to a finite distribution of mass.** Consider a terminating continued fraction of the form  $c_0/1 + g_1 x/1 + (1-g_1)g_2 x/1 + \dots + (1-g_{k-1})g_k x/1$ , where  $c_0 > 0$ ,  $0 < g_n < 1$ ,  $n=1, 2, 3, \dots$ ,  $k-1$ ,  $0 < g_k \leq 1$ . We know that if  $k=2m-1$  this continued fraction represents a rational function of the form

$$\frac{A_{2m}(x)}{B_{2m}(x)} = \frac{M_1}{1+xx_1} + \frac{M_2}{1+xx_2} + \dots + \frac{M_m}{1+xx_m},$$

where  $M_i > 0$ ,  $i=1, 2, 3, \dots, m$ , and  $0 < x_1 < x_2 < \dots < x_m \leq 1$ . On the other hand, if  $k=2m$ , the function represented has the form

$$\frac{A_{2m+1}(x)}{B_{2m+1}(x)} = M_0 + \frac{M_1}{1+xx_1} + \frac{M_2}{1+xx_2} + \dots + \frac{M_m}{1+xx_m},$$

where  $M_i > 0$ ,  $i=0, 1, 2, \dots, m$ , and  $0 < x_1 < x_2 < \dots < x_m \leq 1$ . Naturally the  $M_i$ 's,  $x_i$ 's are not the same in the two cases. In either case if  $c_0 - c_1 x + c_2 x^2 - \dots$  is the corresponding power series, then

$$c_n = \sum_{i=0}^{n-m} x_i^n M_i, \quad n = 0, 1, 2, \dots,$$

where  $x_0=0$ ,  $x_0^0=1$ , and  $M_0$  is positive or 0 according as  $k=2m$  or  $2m-1$ . Thus we may write  $c_n = \int_0^1 u^n d\phi(u)$ ,  $n=0, 1, 2, \dots$ , where  $\phi(u)$  is a monotone nondecreasing function with but a finite number of points of increase.

Conversely, let  $c_0, c_1, c_2, \dots$  be a totally monotone sequence corresponding to a finite distribution of mass. Then  $c_n$  has a representation of the form

$$(6.1) \quad c_n = \int_0^1 u^n d\phi(u), \quad n = 0, 1, 2, \dots,$$

where  $\phi(u)$  is a step-function. There are two cases to be considered according as  $\phi(u)$  is or is not continuous at  $u=0$ .

LEMMA 6.1. Let  $A_n, B_n$  denote the determinants  $|c_{i+j}|, |c_{i+j+1}|$  ( $i, j=0, 1, \dots, n$ ), respectively. Then if  $\phi(u)$  is continuous at  $u=0$ , these determinants are positive for  $n < m$ , and if  $\phi(u)$  is discontinuous at  $u=0$ ,  $A_n > 0$  for  $n \leq m$ , and  $B_n > 0$  for  $n < m$ , where  $m$  is the number of values of  $u > 0$  where  $\phi(u)$  is discontinuous.

**Proof.** Put

$$c_n = \sum_{i=0}^{i=m} x_i M_i, \quad n = 0, 1, 2, \dots,$$

where  $0 < x_1 < x_2 < \dots < x_m \leq 1$ ,  $M_1, M_2, \dots, M_m$  are positive, and  $x_0 = 0$ ,  $x_0^0 = 1$ . If  $\phi(u)$  is continuous at  $u=0$ , then  $M_0 = 0$ , while if  $\phi(u)$  is discontinuous at  $u=0$ ,  $M_0 > 0$ . Consider the quadratic form

$$\sum_{i,j=0}^n c_{i+j+r} \xi_i \xi_j = \int_0^1 u^r \left\{ \sum_{i=0}^{i=n} \xi_i u^i \right\}^2 d\phi(u).$$

If  $r=1$  this is clearly positive definite for  $n < m$ , so that  $B_n > 0$  for  $n < m$  in both cases. When  $r=0$ , it is positive definite for  $n < m$  if  $\phi(u)$  is continuous at  $u=0$ , and for  $n \leq m$  if  $\phi(u)$  is discontinuous at  $u=0$ . Hence  $A_n > 0$  in the first case for  $n < m$ , and in the second for  $n \leq m$ .

Now

$$P(x) \equiv c_0 - c_1 x + c_2 x^2 - \dots \equiv M_0 + \frac{M_1}{1 + x x_1} + \frac{M_2}{1 + x x_2} + \dots \\ + \frac{M_m}{1 + x x_m} = \frac{S(x)}{T(x)},$$

where  $T(x)$  is a polynomial of degree  $m$ , and  $S(x)$  is of degree  $m$  or  $m-1$  according as  $M_0 > 0$  or  $M_0 = 0$ . Put

$$a_1 = c_0, \quad a_{2n} = B_{n-1} A_{n-2} / A_{n-1} B_{n-2}, \quad a_{2n+1} = A_n B_{n-2} / B_{n-1} A_{n-1},$$

$n = 1, 2, 3, \dots$ ,  $A_{-1} = B_{-1} = 1$ , and form the continued fractions

$$(6.2) \quad \frac{A_{2m}(x)}{B_{2m}(x)} = \frac{a_1}{1} + \frac{a_2 x}{1} + \dots + \frac{a_{2m} x}{1} + \frac{0 \cdot x}{1},$$

and

$$(6.3) \quad \frac{A_{2m+1}(x)}{B_{2m+1}(x)} = \frac{a_1}{1} + \frac{a_2 x}{1} + \dots + \frac{a_{2m+1} x}{1} + \frac{0 \cdot x}{1},$$

according as  $M_0 = 0$  or  $M_0 > 0$ . The vanishing partial quotient is affixed for a

reason to appear presently. On taking account of the degrees of numerators and denominators, and of the degree of approximation of these rational fractions to the power series  $P(x)$ , we conclude that they are identical with  $S(x)/T(x)$  in the two cases.

The denominators of the  $n$ th approximants of (6.2), (6.3) are greater than 0 if  $-1 \leq x \leq 0$ , and  $n < 2m$ ,  $n < 2m+1$ , respectively. Hence we may apply Theorem 2.3 with  $c=1$  to show that the  $a_n$ 's have the form  $g_{n-1}(1-g_{n-2})$ ,  $n > 2$ ,  $a_2 = g_1$ . We then apply Theorem 2.2 and obtain in the case of (6.2):

$$\Delta c_0 - \Delta c_1 x + \Delta c_2 x^2 - \dots \sim \frac{\Delta c_0}{1} + \frac{g_1(1-g_1)x}{1} + \dots + \frac{g_{2m-2}(1-g_{2m-1})x}{1} + \frac{g_{2m-1}x}{1}.$$

It is easy to see that  $\phi(u)$  is discontinuous at  $u=1$  if and only if  $g_{2m-1}=1$ . In case  $g_{2m-1}=1$ , the above continued fraction terminates with the  $(2m-2)$ th partial quotient, while if  $g_{2m-1} < 1$  it terminates with the  $2m$ th partial quotient.

Now  $\Delta c_n = \int_0^1 u^n d\phi_1(u)$ , where  $\phi_1(u) = \int_0^u (1-u)d\phi(u)$ . Hence  $\phi_1(u)$  has the same number of discontinuities as, or a smaller number by one than,  $\phi(u)$ , according as 1 is not or is, respectively, a point of discontinuity of  $\phi(u)$ . Since  $\phi_1(u)$  is a function of the same character as  $\phi(u)$ , we conclude that  $1-g_1 > 0$ ,  $g_{n-1}(1-g_n) > 0$ ,  $n=2, 3, \dots, 2m-2$  if  $g_{2m-1}=1$ , and  $n=2, 3, \dots, 2m-1$  if  $g_{2m-1} < 1$ . But we previously had  $g_1 > 0$ ,  $g_n(1-g_{n-1}) > 0$ ,  $n=2, 3, \dots, 2m-1$ . Hence we conclude that  $0 < g_n < 1$ ,  $n=1, 2, 3, \dots, 2m-2$ ,  $0 < g_{2m-1} \leq 1$ . The treatment of (6.3) is exactly the same. We have therefore completed the proof of the following theorem:

**THEOREM 6.1.** *If  $c_0, c_1, c_2, \dots$  is a totally monotone sequence corresponding to a finite distribution of mass, then  $c_0 - c_1x + c_2x^2 - \dots$  is the constant  $c_0 \geq 0$ , or else*

$$c_0 - c_1x + c_2x^2 - \dots \sim c_0/1 + g_1x/1 + (1-g_1)g_2x/1 + \dots + g_k(1-g_{k-1})x/1,$$

where  $0 < g_n < 1$ ,  $n < k$ ,  $0 < g_k \leq 1$ ,  $c_0 > 0$ . Conversely, any sequence determined in this way is a totally monotone sequence corresponding to a finite distribution of mass.

**7. The moment problem for the interval  $(-\infty, 1)$ .** The methods used previously may be employed to prove the following theorem:

**THEOREM 7.1.** *If  $\phi(u)$  is a monotone nondecreasing function in the interval  $(-\infty, 1)$ , such that the moments  $c_n = \int_{-\infty}^1 u^n d\phi(u)$ ,  $n=0, 1, 2, \dots$ , are all finite, and if the series  $c_0 - c_1x + c_2x^2 - \dots$  has a corresponding continued fraction,*

then this continued fraction must have the form

$$(7.1) \quad c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots,$$

where  $c_0 > 0$ , and

$$g_{2n-1}g_{2n} > 0, \quad (1 - g_{2n-1})(1 - g_{2n-2}) > 0,$$

$$n = 1, 2, 3, \dots, g_0 = 0.$$

**Proof.** Let  $b_1/1 + b_2x/1 + b_3x/1 + \dots$  be the corresponding continued fraction which is supposed to exist. Then it is known<sup>(13)</sup> that  $b_1 > 0$ ,  $b_{2n}b_{2n+1} > 0$ ,  $n = 1, 2, 3, \dots$ . The zeros of the denominators  $B_n(x)$  of the approximants of this continued fraction are all real and none of them lie in the interval<sup>(14)</sup>  $-1 \leq x \leq 0$ . Hence  $B_n(-1) > 0$ . By Theorem 2.3 we may then write this continued fraction in the form (7.1) where  $g_n \neq 0$ ,  $1, n \geq 1$ . Then by Theorem 2.2 we have

$$\Delta c_0 - \Delta c_1x + \Delta c_2x^2 - \dots \sim \frac{\Delta c_0}{1} + \frac{g_1(1 - g_2)x}{1} + \frac{g_2(1 - g_3)x}{1} + \dots.$$

Now  $\Delta^m c_n = \int_{-\infty}^1 (1-u)^m u^n d\phi(u) = \int_{-\infty}^1 u^n d\phi_1(u)$ , where  $\phi_1(u) = \int_{-\infty}^u (1-u)^m d\phi(u)$  is monotone nondecreasing for  $-\infty < u \leq 1$ , and consequently  $\Delta c_0 = c_0(1 - g_1) > 0$ ,  $g_{2n-1}g_{2n}(1 - g_{2n})(1 - g_{2n+1}) > 0$ ,  $n \geq 1$ . The theorem now follows. We shall leave unanswered the question of the converse of this theorem.

**8. Moment generating functions which are bounded in the unit circle.** The problem of this section is to specialize the continued fractions of Theorems 4.1 and 6.1 in such a way that the *moment generating function*  $f(x) = c_0 - c_1x + c_2x^2 - \dots$  will be bounded in the unit circle, that is

$$M(f) = \text{l.u.b.}_{|x| < 1} |f(x)|$$

will be finite. Our first result is contained in the theorem which follows.

**THEOREM 8.1.** *The function  $f(x) = c_0 - c_1x + c_2x^2 - \dots$  is in the class E, of moment generating functions bounded in the unit circle, if and only if there is a correspondence of the form*

$$hf(x) \sim g_1/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots,$$

for some sufficiently small positive number  $h$  and for real  $g_n$ 's such that  $0 \leq g_n \leq 1$ ,  $n = 1, 2, 3, \dots$ , where it is agreed that the continued fraction shall terminate with the first identically vanishing partial quotient. When the condition is satisfied,  $M(f) \leq 1/h$ , and  $h$  can be taken equal to 1 if and only if  $M(f) \leq 1$ .

**Proof of sufficiency.** Suppose first that the continued fraction terminates

<sup>(13)</sup> H. Hamburger, *Mathematische Annalen*, vol. 81 (1920), pp. 235-319, and vol. 82 (1921), pp. 120-137.

<sup>(14)</sup> Cf. footnote 8.

and that  $hf(x) = g_1$ ,  $0 \leq g_1 \leq 1$ , or  $hf(x) = g_1/1 + (1-g_1)g_2x/1 + \dots + (1-g_k) \cdot g_{k+1}x/1$ ,  $k \geq 1$ ,  $0 < g_n < 1$ ,  $n = 1, 2, 3, \dots, k$ ,  $0 < g_{k+1} \leq 1$ . The first possibility is at once disposed of. When  $g_{k+1} < 1$ , the second possibility is disposed of by Theorem 3.1; and when  $g_{k+1} = 1$ , it is required to be shown that the rational function  $hf(x)$  has no poles for  $|x| \leq 1$ . To do this, recall that the function  $1/(1+xhf(x))$  has the form

$$\frac{M_0}{1+xx_0} + \frac{M_1}{1+xx_1} + \dots + \frac{M_{m-1}}{1+xx_{m-1}} + \frac{M_m}{1+xx_m}$$

where  $x_0 = 0 < x_1 < x_2 < \dots < x_m \leq 1$ ,  $M_0 \geq 0$ ,  $M_i > 0$ ,  $i = 1, 2, \dots, m$ , and that in the case under consideration  $x_m = 1$ . Consequently, the zero of this function which lies nearest the origin is less than  $-1$ , so that the function  $1+xhf(x)$  has no pole for  $|x| \leq 1$  and therefore  $M(f)$  is finite. It will be seen that  $f(-1) = 1/h = c_0 + c_1 + c_2 + \dots = M(f)$ .

When the continued fraction does not terminate, then by Theorem 3.1,  $M(f) \leq 1/h$ , and  $f(x) \in E$ .

**Proof of necessity.** Suppose conversely that  $f(x) \in E$ , and let  $f(x) = c_0/1 + h_1x/1 + (1-h_1)h_2x/1 + (1-h_2)h_3x/1 + \dots$ , where  $c_0 \geq 0$ ,  $0 \leq h_n \leq 1$ ,  $n \geq 1$ . If  $c_0 = 0$  or  $h_1 = 0$ , the theorem is obviously true. Suppose  $c_0 > 0$ ,  $0 < h_n < 1$ ,  $n = 1, 2, \dots, k-1$ ,  $0 < h_k \leq 1$ , and that the continued fraction terminates by having  $h_{k+1}(1-h_k) = 0$ . Then we must have  $h_k < 1$ , for otherwise  $f(x)$  has a pole at  $x = -1$ . Now consider, whether or not the continued fraction terminates, the function

$$\begin{aligned} 1/(1+xhf(x)) &= 1/1 + hc_0x/1 + h_1x/1 + (1-h_1)h_2x/1 \\ &\quad + (1-h_2)h_3x/1 + \dots \end{aligned}$$

Since  $0 < h_1 < 1$ ,  $0 \leq h_n < 1$ ,  $n > 1$ , the continued fraction  $1 + h_1x/1 + (1-h_1)h_2x/1 + (1-h_2)h_3x/1 + \dots$  converges uniformly for  $|x| \leq 1$ , by Theorem 3.1, and cannot vanish for  $|x| \leq 1$  since  $M(f) < \infty$  by hypothesis. It readily follows that  $h > 0$  can be so chosen ( $h = 1$  if  $M(f) \leq 1$ ), that for every  $r$ ,  $0 < r < 1$ , the continued fraction for  $1/(1+xhf(x))$  converges uniformly for  $|x| \leq r$ , and consequently we must have, for the chosen value of  $h$ ,

$$1/(1+xhf(x)) = \int_0^1 d\phi(u)/(1+xu),$$

where  $\phi(u)$  is some monotone nondecreasing function. Hence  $1/(1+xhf(x))$  is a moment generating function, and therefore has a representation of the form  $1/1 + g_1x/1 + (1-g_1)g_2x/1 + (1-g_2)g_3x/1 + \dots$ , so that  $hf(x)$  has a representation as prescribed by the theorem.

Another characterization of  $E$  in terms of continued fractions is given by the theorem which follows:

**THEOREM 8.2.** *Given a moment generating function  $f(x) = c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + \dots$  ( $c_0 \geq 0$ ,  $0 \leq g_n \leq 1$ ),  $f(x) \in E$  if and only if  $f(x) \equiv c_0 \geq 0$ , or  $f(x) = c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + \dots + (1 - g_{k-2})g_{k-1}x/1$ , where  $c_0 > 0$ ,  $0 < g_n < 1$ ,  $n = 1, 2, 3, \dots, k-1$ ; or else  $f(x) = c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$ , where  $c_0 > 0$ ,  $0 < g_n < 1$ ,  $n \geq 1$ , and the series*

$$(8.1) \quad 1 + \sum_{n=1}^{\infty} \frac{g_1 g_2 \cdots g_n}{(1 - g_1)(1 - g_2) \cdots (1 - g_n)}$$

*converges.*

**Proof.** The case of the terminating continued fraction is easily disposed of, for when the condition is satisfied the rational function has no poles for  $|x| \leq 1$ , while if  $g_{k-1} = 1$  (the only alternative) there is a pole at  $x = -1$ .

In the case of the nonterminating continued fraction, the function  $c_0/f(x) = 1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$  is analytic for  $|x| < 1$  and continuous for  $|x| \leq 1$ . Also by Theorem 2.1 the function  $(1+x)f(x)$  enjoys these same properties. Thus  $f(x)$  is continuous for  $|x| \leq 1$  except possibly at  $x = -1$ . But by Theorem 3.1  $f(-1)/c_0$  is equal to the series (8.1) and is therefore finite if and only if the latter converges.

**9. A characterization of  $E$  in terms of the moments  $c_n$ .** We now prove

**THEOREM 9.1.** *A moment generating function  $f(x) = c_0 - c_1x + c_2x^2 - \dots$  is in  $E$  if and only if the series  $c_0 + c_1 + c_2 + \dots$  converges, and when the condition is satisfied we have*

$$M(f) = c_0 + c_1 + c_2 + \dots$$

**Proof. Sufficiency:** When the series  $c_0 + c_1 + c_2 + \dots$  converges, the sequence  $d_n = c_n + c_{n+1} + c_{n+2} + \dots$ ,  $n = 0, 1, 2, \dots$ , is totally monotone, so that the power series  $d_0 - d_1x + d_2x^2 - \dots$  has a corresponding continued fraction of the form  $d_0/1 + h_1x/1 + (1 - h_1)h_2x/1 + (1 - h_2)h_3x/1 + \dots$ , ( $d_0 \geq 0$ ,  $0 \leq h_n \leq 1$ ). Then, since  $\Delta d_n = c_n$ , it follows from Theorem 2.2 that  $c_0 - c_1x + c_2x^2 - \dots \sim c_0/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$  where we have put  $g_n = 1 - h_n$ ,  $n = 1, 2, 3, \dots$ . That  $f(x) \in E$  now follows from Theorem 8.1.

**Necessity:** The necessity of the condition results at once from a theorem of Abel. However, it is interesting to give a direct proof. Suppose then that  $f(x) \in E$ . Then by Theorem 8.1 we must have  $hf(x) \sim g_1/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$  for some  $h > 0$ , where the  $g_n$ 's are restricted as in Theorem 8.1. Therefore  $1 + hx f(x) \sim 1 + g_1x + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$ , so that, by Theorem 2.1,

$$\frac{1 + hx f(x)}{1 + x} \sim \frac{1}{1 + \frac{(1 - g_1)x}{1} + \frac{g_1(1 - g_2)x}{1} + \dots}$$

Consequently the function  $(1 + hx f(x))/(1 + x)$  is a moment generating func-



tion. Its power series expansion is  $1 - (1 - hc_0)x + (1 - hc_0 - hc_1)x^2 - \dots$ , so that  $1 - h(c_0 + c_1 + c_2 + \dots + c_n) \geq 0$  for  $n = 0, 1, 2, \dots$ . It follows that the series  $\sum c_i$  is convergent.

Since  $f(-1) = c_0 + c_1 + c_2 + \dots$ , it follows that  $M(f)$  is equal to this sum.

As a corollary to Theorem 9.1 we have

**THEOREM 9.2.** *The moment generating function  $f(x)$  is in  $E$  if and only if it has a Stieltjes integral representation of the form*

$$f(x) = \int_0^1 \frac{(1-u)d\phi(u)}{1+xu},$$

in which  $\phi(u)$  is bounded and monotone nondecreasing.

**Proof.** If  $f(x) \in E$ , then the series  $\sum c_i$  converges. Let  $d_n$  be defined as in the proof of Theorem 9.1. Then the function  $d_0 - d_1x + d_2x^2 - \dots$  has a representation of the form  $\int_0^1 d\phi(u)/(1+xu)$ , where  $\phi(u)$  is bounded and monotone nondecreasing. Since we must have  $f(x) = \int_0^1 (1-u)d\phi(u)/(1+xu)$ , the necessity of the condition is proved. Conversely, if the condition is fulfilled, put  $d_0 - d_1x + d_2x^2 - \dots = \int_0^1 d\phi(u)/(1+xu)$ , where  $\phi(u)$  is given. Then  $c_n = \Delta d_n$ , so that  $\sum c_i = d_0 - \lim_{n \rightarrow \infty} d_n$  is convergent, and consequently  $f(x) \in E$ .

**10. The algorithm of Schur.** Starting with a function  $f(x)$  of  $E$  for which  $M(f) \leq 1$ , we construct with Schur<sup>(15)</sup> a sequence of functions  $f_n(x)$  in the following way. Put  $f_0(x) = f(x)$ ,

$$(10.1) \quad f_{n+1}(x) = \frac{1}{x} \frac{t_n - f_n(x)}{1 - t_n f_n(x)}, \quad t_n = f_n(0), \quad n = 0, 1, 2, \dots$$

If  $M(f) = 1$ , then  $M(f_n) = 1$ , and if  $M(f) < 1$ , then  $M(f_n) < 1$ , ( $n = 1, 2, 3, \dots$ ).

If  $f(x) \equiv g_1$ , then  $f_n(x) \equiv 0$  for  $n = 1, 2, 3, \dots$ . If  $f(x) = g_1/1 + (1 - g_1)g_2x/1$ , then  $f_1(x) = g_1^{(1)}/1 + (1 - g_1^{(1)})g_2^{(1)}x/1$ , where

$$g_1^{(1)} = g_1g_2/(1 + g_1), \quad g_2^{(1)} = g_1g_2/(1 + g_1 - g_1g_2),$$

so that if  $0 < g_1 < 1$ ,  $0 < g_2 \leq 1$ , then  $0 < g_1^{(1)} < 1$ ,  $0 < g_2^{(1)} \leq 1$ . We shall prove this theorem:

**THEOREM 10.1.** *If  $f(x) = g_1/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$ ,  $0 \leq g_n \leq 1$ ,  $n = 1, 2, 3, \dots$ , with the agreement that the continued fraction shall terminate with the first identically vanishing partial numerator, so that  $f(x)$  is a moment generating function and  $M(f) \leq 1$ , then the functions  $f_1(x), f_2(x), f_3(x), \dots$  given by (10.1) are all moment generating functions and  $M(f_n) \leq 1$ . The only case where any of these functions are constants is where  $f(x) \equiv g_1$ , whereupon  $f_n(x) \equiv 0$ ,  $n = 1, 2, 3, \dots$ .*

**Proof.** By (10.1), if  $f(x)$  is not a constant, then

<sup>(15)</sup> Cf. footnote 4.

$$(10.2) \quad xf_1(x) = g_1 - \frac{(1 - g_1^2)}{-g_1 + 1/f(x)} \\ = g_1 - g_1/1 + g_2(1 - h_1)x/1 + g_3(1 - g_2)x/1 + \dots,$$

where we have put  $1 - h_1 = 1/(1 + g_1)$ ,  $h_1 = g_1/(1 + g_1)$ . This function is evidently of the form  $g_1 - g(x)$ , where  $g(x)$  is a moment generating function and  $g(0) = g_1$ . Put  $g(x) = g_1 - d_0x + d_1x^2 - \dots$ , and we find that  $f_1(x) = d_0 - d_1x + d_2x^2 - \dots$  and is therefore a moment generating function. On account of the character of the transformation used,  $M(f_1) \leq 1$ . Consequently, by Theorem 8.1,  $f_1(x) = g_1^{(1)}/1 + (1 - g_1^{(1)})g_2^{(1)}x/1 + (1 - g_2^{(1)})g_3^{(1)}x/1 + \dots$  where  $0 \leq g_n^{(1)} \leq 1$ ,  $n = 1, 2, 3, \dots$ , with the oft repeated convention regarding termination of the continued fraction.

In the same way, starting with  $f_1(x)$  instead of with  $f(x)$ , we find that  $f_2(x)$  is a moment generating function, and  $M(f_2) \leq 1$ ; and by induction,  $f_3(x), f_4(x), \dots$  all have this property.

We saw previously that when  $f(x) = g_1$ , then  $f_n(x) = 0$ , ( $n = 1, 2, 3, \dots$ ). It remains to be shown that this is the only case where any of the functions can reduce to a constant. To do this, it suffices to show that if  $f_1(x) \equiv c$ , a constant, then  $c = 0$ . We have

$$f(x) = \frac{g_1 - cx}{1 - cxg_1} = g_1 - (1 - g_1^2)cx - g_1c^2(1 - g_1^2)x^2 - \dots$$

Since this is a moment generating function, we must have  $-g_1c^2(1 - g_1^2) \geq 0$ , which implies that  $c = 0$ , or else  $g_1 = 0$  or  $1$ . But when  $g_1 = 0$  or  $1$ , we must have  $f = g_1$ , and consequently  $f_1 = c = 0$  in this case also.

As a corollary we have

**THEOREM 10.2.** *If  $f(x)$  is a moment generating function for which  $M(f) \leq 1$ , then the sequence  $t_0, t_1, t_2, \dots$  given by (10.1) has the property of a totally monotone sequence that if any member is 0 the others are also 0 with the possible exception of the first.*

In a number of examples which we have examined, the sequence  $\{t_n\}$  has been found to be totally monotone. Technical difficulties have thus far prevented us from determining whether or not this is always the case. Also, the question as to the converse of this naturally arises, namely, if  $\{t_n\}$  is a totally monotone sequence, will the function  $f(x)$  which this sequence determines be a moment generating function with  $M(f) \leq 1$ ?

In conclusion we shall give recursion formulas for computing the  $t_n$ 's in terms of the  $g_n$ 's. From (10.2) we have

$$xf_1(x) = g_1 - \frac{g_1}{1} + \frac{(1 - h_1)g_2x}{1} + \frac{(1 - g_2)g_3x}{1} + \dots,$$

which is equal to  $g_1^{(1)}x/1 + (1 - g_1^{(1)})g_2^{(1)}x/1 + (1 - g_2^{(1)})g_3^{(1)}x/1 + \dots$ . On equating the odd part of the first of these continued fractions to the even part of the second we obtain at once the formulas

$$(10.3) \quad \begin{aligned} g_1^{(1)} &= g_1 g_2 / (1 + g_1), & g_2^{(1)} (1 - g_1^{(1)}) &= \frac{g_2}{1 + g_1} + g_3 (1 - g_2), \\ g_n^{(1)} g_{n+1}^{(1)} (1 - g_{n-1}^{(1)}) (1 - g_n^{(1)}) &= g_{n+1} g_{n+2} (1 - g_n) (1 - g_{n+1}), \\ g_{n+1}^{(1)} (1 - g_n^{(1)}) + g_{n+2}^{(1)} (1 - g_{n+1}^{(1)}) &= g_{n+2} (1 - g_{n+1}) + g_{n+3} (1 - g_{n+2}), \end{aligned}$$

$n=2, 4, 6, \dots$ . By means of these relations one may show that the sequence  $t_n = g_n^{(n)}$ ,  $n=0, 1, 2, \dots$  ( $g_1^{(0)} = g_1$ ) is monotone decreasing. In fact,  $t_n < t_{n-1} / (1 + t_{n-1})$ ,  $n=1, 2, 3, \dots$ .

Using a result of Schur<sup>(16)</sup> one may obtain formulas for the  $t_n$ 's in terms of the  $g_n$ 's. To do this, let the  $n$ th approximant of the continued fraction for  $f(x)$  be

$$\frac{A_n(x)}{B_n(x)} = \frac{a_0 + a_1 x + a_2 x^2 + \dots + a_k x^k}{b_0 + b_1 x + b_2 x^2 + \dots + b_r x^r}.$$

This rational function is a moment generating function, and its modulus is less than or equal to 1 for  $|x| \leq 1$  provided  $M(f) \leq 1$ . Moreover, the sequence  $\{t_i\}$  for this function will be in agreement with that for  $f(x)$  up to and including the term of index  $n-1$ . Moreover, the  $a_i$ 's and  $b_i$ 's can be computed in terms of the  $g_i$ 's. Using the notation of Schur, we then put

$$\delta_m = \begin{vmatrix} b_0 & 0 & \dots & 0 & a_0 & a_1 & \dots & a_{m-2} & a_{m-1} \\ b_1 & b_0 & \dots & 0 & 0 & a_0 & \dots & a_{m-3} & a_{m-2} \\ \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ b_{m-1} & b_{m-2} & \dots & b_0 & 0 & 0 & \dots & 0 & a_0 \\ a_0 & 0 & \dots & 0 & b_0 & b_1 & \dots & b_{m-2} & b_{m-1} \\ a_1 & a_0 & \dots & 0 & 0 & b_0 & \dots & b_{m-3} & b_{m-2} \\ \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ a_{m-1} & a_{m-2} & \dots & a_0 & 0 & 0 & \dots & 0 & b_0 \end{vmatrix}$$

where  $a_i = 0$  if  $i > k$  and  $b_i = 0$  if  $i > r$ . A formula of Schur<sup>(17)</sup> then gives for  $t_i$

$$t_i = \left\{ 1 - \frac{\delta_{i-1} \delta_{i+1}}{\delta_i^2} \right\}^{1/2},$$

which is valid for  $i=0, 1, 2, \dots, n-1$ .

<sup>(16)</sup> Schur, loc. cit. (first part), pp. 213-215.

<sup>(17)</sup> Loc. cit., p. 215, formula (12).

# HAUSDORFF METHODS OF SUMMATION AND CONTINUED FRACTIONS

BY

H. L. GARABEDIAN AND H. S. WALL

1. **Introduction.** We shall be occupied in this paper with a special study of the transformation

$$(1.1) \quad t_m = \sum_{n=0}^m a_{mn} s_n, \quad m, n = 0, 1, 2, \dots,$$

where  $\{s_n\}$  is an infinite sequence of numbers. If, in particular, we are concerned with the infinite series  $\sum_{r=0}^{\infty} u_r$ , then  $s_n = \sum_{r=0}^n u_r$ . Let  $\mathfrak{A} = (a_{mn})$  represent the triangular matrix of the transformation (1.1). Then, the sequence  $\{t_m\}$  is called the *transform* of  $\{s_n\}$  by the matrix  $\mathfrak{A}$  and is represented in symbolic form by  $t_m = \mathfrak{A}\{s_n\}$ . If the matrix  $\mathfrak{A}$  is *regular*, in the sense of Silverman [1] and Toeplitz [2], then the matrix  $\mathfrak{A}$  defines a regular method of summation (*summability*).

In this paper our attention is focused on a class of regular and permutable matrices  $\{\mathfrak{A}\}$ , known as *Hausdorff matrices* ([3] or [4]), which we shall presently define. Let  $\{c_m\}$  be an infinite sequence of numbers defined by the Stieltjes integrals

$$(1.2) \quad c_m = \int_0^1 u^m d\phi(u), \quad m = 0, 1, 2, \dots$$

Then, we form the matrix  $\mathfrak{D}(\delta_{mn}c_m)\mathfrak{D}$ , where  $\delta_{mn}=0$  for  $m \neq n$ ,  $\delta_{nn}=1$ , and  $\mathfrak{D} = ((-1)^n c_{m,n})$ . Incidentally, the matrix  $\mathfrak{D}$  has the property  $\mathfrak{D}^2 = \mathfrak{I}$ , where  $\mathfrak{I}$  is the identity matrix. The following conditions on the *mass function*  $\phi(u)$  are necessary and sufficient in order that the matrix  $\mathfrak{D}(\delta_{mn}c_m)\mathfrak{D}$  be regular:

- (1.3) (i)  $\phi(u)$  is of bounded variation on the closed interval  $(0, 1)$ ;  
(ii)  $\phi(u)$  is continuous at  $u=0$  and  $\phi(u)=0$ ;  
(iii)  $\phi(1)=1$ ;  
(iv)  $\phi(u) = \frac{1}{2}[\phi(u-0) + \phi(u+0)]$ ,  $0 < u < 1$ .

If  $\phi(u)$  satisfies the conditions (1.3), then the matrix  $\mathfrak{D}(\delta_{mn}c_m)\mathfrak{D}$  is a Hausdorff matrix and the sequence (1.2) is known as a *regular sequence* or a *regular moment sequence*. If  $\mathfrak{A} = \mathfrak{D}(\delta_{mn}c_m)\mathfrak{D}$  Hausdorff has proved that we can write

$$(1.4) \quad a_{mn} = C_{m,n} \Delta^{m-n} c_n.$$

Presented to the Society, April 13, 1940; received by the editors February 2, 1940, and, in revised form, February 24, 1940.

We shall be concerned for the most part with real *totally monotone* sequences  $\{c_n\}$  which are characterized in any one of the following three ways:

(1.5) (i)  $\Delta^m c_n \geq 0$ , ( $m, n = 0, 1, 2, \dots$ );

(ii) there exists (essentially uniquely<sup>(1)</sup>) a real monotone non-decreasing  $\phi(u)$  such that  $c_n = \int_0^1 u^n d\phi(u)$ , ( $n = 0, 1, 2, \dots$ );

(iii) there is a correspondence<sup>(2)</sup> of the form

$$c_0 - c_1x + c_2x^2 - \dots \sim \frac{c_0}{1} + \frac{g_1x}{1} + \frac{(1-g_1)g_2x}{1} + \frac{(1-g_2)g_3x}{1} + \dots,$$

where  $c_0 \geq 0$ ,  $0 \leq g_n \leq 1$ , ( $n = 1, 2, 3, \dots$ ), and where it is understood that the continued fraction shall terminate with the first partial quotient which vanishes identically.

These characterizations are due respectively to I. Schur, F. Hausdorff [5], and H. S. Wall [6].

In §2 of this paper we give necessary and sufficient conditions for the regularity of a totally monotone sequence in terms of the corresponding continued fraction. In §3 we investigate certain properties of the *difference matrix*  $\Delta = (\Delta^m c_n)$ . Any row, column, or diagonal sequence of the difference matrix is found to be totally monotone. We obtain necessary and sufficient conditions, in terms of the continued fraction corresponding to the *base sequence*, that is, the sequence defined by the first row of  $\Delta$ , in order that the row, column, and diagonal sequences be regular. This is accomplished with the aid of the following curious result: if (1.5, iii) obtains, then the power series  $c_0 - \Delta c_0x + \Delta^2 c_0x^2 - \dots$ , with coefficients in the first column of the difference matrix  $\Delta$ , corresponds to the continued fraction

$$(1.6) \quad \frac{c_0}{1} + \frac{(1-g_1)x}{1} + \frac{g_1g_2x}{1} + \frac{(1-g_2)(1-g_3)x}{1} + \frac{g_2g_3x}{1} + \dots,$$

obtained from the continued fraction of (1.5, iii) by replacing  $g_{2n-1}$  by  $1 - g_{2n-1}$ , ( $n = 1, 2, 3, \dots$ ). Section 4 is devoted to an example which illustrates some of the results of the two preceding sections.

The remainder of this paper has to do with special Hausdorff methods of summation. In §5 the continued fraction of Gauss is employed to obtain a regular sequence which results in the definition of *hypergeometric summability*. Numerous inclusion and equivalence relations between the hypergeometric methods are derived. In §6 we replace the base sequence of  $\Delta$  by known se-

<sup>(1)</sup> This means that  $\phi(u)$  exists uniquely except for an additive constant at all points of continuity.

<sup>(2)</sup> We use the symbol  $\sim$  between a power series and a continued fraction to indicate that the power series expansion of the  $n$ th approximant of the continued fraction agrees term by term with the given power series for more and more terms as  $n$  is increased, or becomes identical with it from and after some  $n$ .

quences and discuss the new methods of summability thus generated and some of their properties. Finally, in §7, we discuss the effectiveness of methods of summation associated with regular totally monotone sequences relative to the analytic continuation of power series outside the circle of convergence.

2. **The Stieltjes continued fraction as a tool in the theory of summability.** In his celebrated memoir *Recherches sur les fractions continues*, Stieltjes [7] showed that a real sequence  $\{c_n\}$  is a moment sequence for an infinite distribution of mass along the positive half of the real axis if and only if the power series  $c_0 - c_1x + c_2x^2 - \dots$  has a corresponding continued fraction of the form  $b_1/1 + b_2x/1 + b_3x/1 + \dots$ , where the  $b_n$ 's are real and positive. The continued fraction is uniquely determined by the moment sequence, and conversely there is uniquely determined a moment sequence by means of a continued fraction of the specified form. Moreover, the question as to the uniqueness of the distribution of mass for a given moment sequence can always be decided when the continued fraction is known.

When the continued fraction converges, Stieltjes showed that the function represented has the form

$$\int_0^\infty \frac{d\phi(u)}{1+xu},$$

where  $\phi(u)$  is a bounded monotone non-decreasing function, and represents the distribution of mass in accordance with the relations  $c_n = \int_0^\infty u^n d\phi(u)$ , ( $n=0, 1, 2, \dots$ ), the sequence  $\{c_n\}$  defining the given moments. In this case the moment problem is *determinate*, that is, there is but one possible distribution of mass. On the other hand, if the continued fraction diverges, the sequences of even and odd approximants converge to separate integrals of the above form, which determine two distinct solutions of the moment problem. In this, the *indeterminate* case, there are infinitely many distributions of mass for the given moments.

Until recently it was not known how to specialize the numbers  $b_n$  in the continued fraction in the case of moments for a distribution of mass over the finite interval  $(0, 1)$ . The answer to this question is contained in the statement (1.5, iii), where, in the case of the terminating continued fraction, the moments are for a finite distribution of mass.

The moment problem for the interval  $(0, 1)$  was solved without the use of continued fractions by Hausdorff [3]. He found, among other things, that these sequences are regular<sup>(\*)</sup> if and only if the mass function  $\phi(u)$  of (1.5, ii) is continuous at  $u=0$ , and  $\int_0^1 d\phi(u) = c_0 = 1$ . One of the ways in which the con-

(\*) It is sometimes convenient to state the regularity conditions (1.3) in this form. The requirements  $\phi(0)=0$ ,  $\phi(1)=1$  can be replaced by the single condition  $\int_0^1 d\phi(u)=1$ . In the discussion which follows we shall always assume without further mention that this requirement is met. The regularity condition (1.3, iv) is in a sense superfluous since it serves merely to determine  $\phi(u)$  uniquely at every point of the interval  $(0, 1)$ .



tinued fraction may serve as a tool in this theory is in establishing the continuity or the discontinuity of  $\phi(u)$  at  $u=0$ .

We shall begin by disposing of the case where the continued fraction (1.5, iii) terminates, the case corresponding to a finite distribution of mass. Here the function  $\phi(u)$  is a step function with but a finite number of discontinuities, one at each point where a quantity of mass is concentrated. *There is no discontinuity at  $u=0$  if and only if for some index  $n$  the first  $2n$  partial quotients of the continued fraction are not identically zero while the next partial quotient vanishes identically.* This [6] is a consequence of the fact that in this case, and only this, the continued fraction may be written as a sum of partial fractions of the form  $\sum_{i=1}^n M_i/(1+xx_i)$ , without a constant term, where  $M_i > 0$ ,  $0 < x_1 < x_2 < \dots < x_n \leq 1$ . The moments are then  $c_m = \sum_{i=1}^n x_i^m M_i$ , the function  $\phi(u)$  having discontinuities only at the points  $x_i$ .

When the continued fraction does not terminate, it may always be written in the form  $1/a_1 + x/a_2 + x/a_3 + \dots$ , where the  $a_n$ 's are positive. Then from the work of Stieltjes [8, p. 510] we have the theorem that  $\phi(u)$  is continuous at  $u=0$  if and only if the series  $\sum a_{2n-1}$  diverges. Since

$$\begin{aligned} a_1 &= 1/c_0, \quad a_3 = g_1/c_0 g_2(1-g_1), \dots, \\ a_{2n+1} &= [g_1 g_3 \dots g_{2n-1}(1-g_2)(1-g_4) \dots (1-g_{2n-2})] \\ &\quad \div [c_0 g_2 g_4 \dots g_{2n}(1-g_1)(1-g_3) \dots (1-g_{2n-1})], \end{aligned}$$

this condition appears at once in terms of the parameters  $g_n$ . Remembering that in addition to the continuity of  $\phi(u)$  at  $u=0$  we require for regularity that  $c_0=1$ , we have the following theorem.

**THEOREM 2.1.** *The totally monotone sequence  $\{c_n\}$  is regular if and only if the power series  $c_0 - c_1 x + c_2 x^2 - \dots$  has a terminating corresponding continued fraction of the form*

$$(2.1) \quad \frac{1}{1} + \frac{g_1 x}{1} + \frac{(1-g_1)g_2 x}{1} + \dots + \frac{(1-g_{2n-2})g_{2n-1} x}{1},$$

where  $0 < g_k < 1$ , ( $k=1, 2, 3, \dots, 2n-2$ ),  $0 < g_{2n-1} \leq 1$  (in which case the distribution of mass is finite), or else has a nonterminating continued fraction of the form

$$(2.2) \quad \frac{1}{1} + \frac{g_1 x}{1} + \frac{(1-g_1)g_2 x}{1} + \frac{(1-g_2)g_3 x}{1} + \dots,$$

where  $0 < g_n < 1$ , ( $n=1, 2, 3, \dots$ ), and the series

$$(2.3) \quad \sum \frac{g_1 g_3 \dots g_{2n+1}(1-g_2)(1-g_4) \dots (1-g_{2n})}{g_2 g_4 \dots g_{2n+2}(1-g_1)(1-g_3) \dots (1-g_{2n+1})}$$

diverges (in which case there is an infinite distribution of mass).

We shall record here for future reference some known properties of the special Stieltjes continued fraction with which we are concerned. These will be stated in the form of theorems, with adequate references.

**THEOREM 2.2** [6, p. 166]. *If  $g_1, g_2, g_3, \dots$  are any real or complex numbers, and  $P(x)$  is a power series in ascending powers of  $x$  such that*

$$P(x) \sim 1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots,$$

*then*

$$(1 + x)/P(x) \sim 1 + (1 - g_1)x/1 + g_1(1 - g_2)x/1 + g_2(1 - g_3)x/1 + \dots.$$

*If  $c_0 \neq 0$  and we put  $c_0/P(x) = c_0 - c_1x + c_2x^2 - \dots$ , this statement takes the following form: if  $c_0 - c_1x + c_2x^2 - \dots \sim c_0/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$ , then  $\Delta c_0 - \Delta c_1x + \Delta c_2x^2 - \dots \sim \Delta c_0/1 + g_1(1 - g_2)x/1 + g_2(1 - g_3)x/1 + \dots$ .*

**THEOREM 2.3** [9, p. 159]. *If  $g_1, g_2, g_3, \dots$  are real,  $0 < g_1 < 1$ ,  $0 \leq g_n < 1$ ,  $n > 1$ , then the continued fraction*

$$(2.4) \quad g_1/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots$$

*converges uniformly for  $|x| \leq 1$ . The function  $f(x)$  represented is continuous for  $|x| \leq 1$ , analytic for  $|x| < 1$ , and its modulus for  $|x| \leq 1$  does not exceed*

$$1 - \left[ \sum \frac{g_1 g_2 \dots g_n}{(1 - g_1)(1 - g_2) \dots (1 - g_n)} \right]^{-1}.$$

*This is the least upper bound, since it is assumed by  $f(x)$  at  $x = -1$ .*

*The continued fraction converges uniformly over every bounded closed region containing no real point  $x$  which is less than or equal to  $-1$ , and  $f(x)$  is analytic in every such region [8].*

**3. The difference matrix.** Let  $\{c_n\}$  be a totally monotone sequence. Then

$$c_n = \int_0^1 u^n d\phi(u)$$

where  $\phi(u)$  is monotone non-decreasing. Now, the  $m$ th difference of  $c_n$  is given by the formula

$$\Delta^m c_n = \int_0^1 (1 - u)^m u^n d\phi(u), \quad m, n = 0, 1, 2, \dots$$

If we keep  $m$  fixed and allow  $n$  to vary, it is evident that the resulting sequence  $(\Delta^m c_0, \Delta^m c_1, \Delta^m c_2, \dots)$  is totally monotone, the mass function being  $\int_0^1 (1 - t)^m d\phi(t)$ . Moreover, if we keep  $n$  fixed and allow  $m$  to vary we obtain the sequence  $(c_n, \Delta c_n, \Delta^2 c_n, \dots)$ . Inasmuch as  $\Delta^m c_n = \int_0^1 u^n [-(1 - u)^m] d\phi(1 - u)$ , it is clear that this sequence is totally monotone, the mass function being

$\int_0^u [-(1-t)^n] d\phi(1-t)$ . Thus, the row and column sequences of the difference matrix  $\Delta$  are all totally monotone.

Consider next a diagonal sequence

$$(3.1) \quad c_n, \Delta c_{n+1}, \Delta^2 c_{n+2}, \dots,$$

or

$$(3.2) \quad \Delta^n c_0, \Delta^{n+1} c_1, \Delta^{n+2} c_2, \dots$$

We may write

$$\Delta^k c_{n+k} = \int_0^{1/4} u^k d\phi'(u), \quad \Delta^{n+k} c_k = \int_0^{1/4} u^k d\phi''(u),$$

where

$$d\phi'(u) = u_1^n d\phi(u_1) - u_2^n d\phi(u_2), \quad d\phi''(u) = u_2^n d\phi(u_1) - u_1^n d\phi(u_2),$$

and  $u_1$  is the smallest and  $u_2$  the largest of the roots of the quadratic equation  $u^2 - u + v = 0$ ,  $0 \leq v \leq \frac{1}{4}$ . Accordingly we observe that the sequences (3.1) and (3.2) are totally monotone.

An inspection of the mass functions for these row, column, and diagonal sequences reveals that they can all be made regular, by dividing all members of each by its first member, if and only if  $\phi(u)$  is continuous at  $u=0$  and at  $u=1$ . In Theorem 2.1 we gave necessary and sufficient conditions for the continuity at  $u=0$ , in terms of the continued fraction corresponding to the base sequence. We now propose to do likewise for the point  $u=1$ .

The case of a finite distribution of mass is disposed of by means of the next theorem, which follows from the work of Wall [6].

**THEOREM 3.1.** *Let  $\{c_n\}$  be a totally monotone sequence corresponding to a finite distribution of mass, and let  $c_0/1 + g_1x/1 + (1-g_1)g_2x/1 + \dots + (1-g_{n-1})g_nx/1$  be the corresponding (necessarily terminating) continued fraction, where  $c_0 > 0$ ,  $0 < g_i < 1$ , ( $i=1, 2, 3, \dots, n-1$ ),  $0 < g_n \leq 1$ . Then the mass function  $\phi(u)$  is continuous at  $u=1$  if and only if  $g_n < 1$ .*

Turning to the case of an infinite distribution of mass, we consider the power series  $c_0 - \Delta c_0x + \Delta^2 c_0x^2 - \dots$  with coefficients from the first column of the difference matrix. Let  $c_0/1 + h_1x/1 + (1-h_1)h_2x/1 + \dots$  be the corresponding continued fraction. Then it is apparent, in view of Theorem 2.1, that  $\phi(1-u)$  will be continuous at  $u=0$ , that is,  $\phi(u)$  will be continuous at  $u=1$ , if and only if the series obtained from (2.3) by replacing  $g_n$  by  $h_n$ , ( $n=1, 2, 3, \dots$ ), is divergent. The problem will therefore be solved if we determine the  $h_n$ 's as functions of the  $g_n$ 's. We shall prove that  $h_{2n} = g_{2n}$ ,  $h_{2n-1} = 1 - g_{2n-1}$ , ( $n=1, 2, 3, \dots$ ). In fact, the following theorem provides a companion theorem to Theorem 2.2.

THEOREM 3.2. Let  $c_0$  be different from 0, and let  $g_1, g_2, g_3, \dots$  be arbitrary real or complex numbers. Then, if

$$(3.3) \quad c_0 - c_1x + c_2x^2 - \dots \sim \frac{c_0}{1} + \frac{g_1x}{1} + \frac{(1-g_1)g_2x}{1} + \frac{(1-g_2)g_3x}{1} + \dots,$$

we have

$$(3.4) \quad \begin{aligned} &c_0 - \Delta c_0x + \Delta^2 c_0x^2 - \dots \\ &\sim \frac{c_0}{1} + \frac{(1-g_1)x}{1} + \frac{g_1g_2x}{1} + \frac{(1-g_2)(1-g_3)x}{1} + \dots, \end{aligned}$$

where the second continued fraction is obtained from the first by replacing  $g_{2n-1}$  by  $1-g_{2n-1}$ , ( $n=1, 2, 3, \dots$ ).

Let

$$(3.5) \quad \begin{aligned} &c_0 - a_1x + a_2x^2 - \dots \\ &\sim \frac{c_0}{1} + \frac{(1-g_1)x}{1} + \frac{g_1g_2x}{1} + \frac{(1-g_2)(1-g_3)x}{1} + \dots \end{aligned}$$

It is required to prove that  $a_n = \Delta^n c_0$ . Replace  $x$  by  $-x$ , multiply by  $x$ , and then replace  $x$  by  $x/(1+x)$  in the last relation. We obtain

$$(3.6) \quad \begin{aligned} &c_0 \left( \frac{x}{1+x} \right) + a_1 \left( \frac{x}{1+x} \right)^2 + a_2 \left( \frac{x}{1+x} \right)^3 + \dots \\ &\sim \frac{c_0}{1+x} - \frac{(1-g_1)x}{1} - \frac{g_1g_2x}{1+x} - \frac{(1-g_2)(1-g_3)x}{1} - \dots \end{aligned}$$

Now, the even part of the last continued fraction (that is, the continued fraction having as its sequence of approximants the even approximants of this continued fraction) is the same as the even part of the continued fraction of (3.3). It follows that the formal power series expansion of the left-hand member of (3.6) is  $c_0 - c_1x + c_2x^2 - \dots$  and hence that  $a_n$  must equal  $\Delta^n c_0$ .

With reference to the difference matrix  $\Delta$  the principal result of this discussion may conveniently be summarized in the following theorem.

THEOREM 3.3. Let  $\{c_n\}$  be a totally monotone sequence corresponding to an infinite distribution of mass. Then the sequence  $(c_0, \Delta c_0, \Delta^2 c_0, \dots)$ , the first column of  $\Delta$ , is regular if and only if  $c_0 = 1$  and the series

$$(3.7) \quad \sum \frac{(1-g_1)(1-g_2) \dots (1-g_{2n-1})}{g_1g_2 \dots g_{2n}}$$

diverges, where the  $g_n$ 's are related to the  $c_n$ 's as in (1.5, iii). The sequence

$(c_0, \Delta c_1, \Delta^2 c_2, \dots)$ , the principal diagonal of  $\Delta$ , is regular if and only if  $c_0 = 1$ , and both of the series (2.3) and (3.7) are divergent.

We have given conditions for regularity of certain moment sequences, chosen from the difference matrix, in terms of the continued fraction corresponding to the base sequence. The conditions can also be given in terms of the moment generating function

$$f(x) = \int_0^1 \frac{d\phi(u)}{1+xu}.$$

In fact, Schoenberg [10] gave necessary and sufficient conditions, in terms of  $f(x)$ , for the continuity of  $\phi(u)$  at an arbitrary point  $u=t$ ,  $0 \leq t \leq 1$ . We may therefore state the theorem which follows.

**THEOREM 3.4.** *A totally monotone sequence  $\{c_n\}$  with corresponding mass function  $\phi(u)$  is regular if and only if  $c_0 = 1$  and*

$$(3.8) \quad \lim_{x \rightarrow \infty} \int_0^1 \frac{d\phi(u)}{1+xu} = 0,$$

where  $x \rightarrow \infty$  along any ray except the negative half of the real axis. The sequence  $(c_0, \Delta c_0, \Delta^2 c_0, \dots)$  is regular if and only if  $c_0 = 0$ , and

$$(3.9) \quad \lim_{x \rightarrow -1} (1+x) \int_0^1 \frac{d\phi(u)}{1+xu} = 0,$$

where  $x \rightarrow -1$  through values interior to or upon the circle  $|x| = 1$ . Finally, the sequence  $(c_0, \Delta c_1, \Delta^2 c_2, \dots)$  is regular if and only if  $c_0 = 1$ , and (3.8), (3.9) both hold.

The regularity conditions can also be given in terms of the moments themselves. Indeed, it is not difficult to show that  $\lim_{n \rightarrow \infty} c_n = \phi(1) - \phi(1-0)$ , and  $\lim_{n \rightarrow \infty} \Delta^n c_0 = \phi(+0) - \phi(0)$ . Hence we have the following theorem.

**THEOREM 3.5.** *The limits (3.8) and (3.9) in Theorem 3.4 may be replaced by the limits*

$$(3.10) \quad \lim_{n \rightarrow \infty} \Delta^n c_0 = 0,$$

and  $\lim_{n \rightarrow \infty} c_n = 0$ , respectively.

If the limit (3.10) is  $k > 0$ , then  $k$  is the amount of the discontinuity of  $\phi(u)$  at  $u=0$ . If then we subtract  $k$  from  $c_0$  in the sequence  $\{c_n\}$ , the resulting sequence is totally monotone, and can be made regular by dividing every member by  $c_0 - k$ .

Wall [6] considered the class of moment generating functions  $f(x) = \sum_{i=0}^{\infty} c_i (-x)^i$  in which the coefficients  $c_n$  form a totally monotone sequence.

In particular, he characterized in several ways the subclass of these functions which are bounded in the unit circle. It is of interest to apply Theorem 3.2 to obtain results of this kind.

**THEOREM 3.6.** *If  $f(x) = c_0/1 + g_1x/1 + (1-g_1)g_2x/1 + (1-g_2)g_3x/1 + \dots$ , where  $c_0 > 0$ ,  $0 < g_n < 1$ ,  $n \geq 1$ , so that  $f(x)$  is a moment generating function corresponding to an infinite distribution of mass, then the modulus of the function  $(1+x)f(x)$  is bounded in the half-plane  $\Re(x) > -\frac{1}{2}$  if and only if the series*

$$(3.11) \quad 1 + \frac{(1-g_1)}{g_1} + \frac{(1-g_1)g_2}{g_1(1-g_2)} + \frac{(1-g_1)g_2(1-g_3)}{g_1(1-g_2)g_3} + \dots$$

is convergent.

If we put  $f_1(x) = c_0 - \Delta c_0x + \Delta^2 c_0x^2 - \dots$ , then  $(1+x)f(x) = f_1(w)$  where  $w = -x/(1+x)$ . Now  $\Re(x) > -\frac{1}{2}$  if and only if  $|w| < 1$ . But [6, p. 181] if  $f_1(w) = c_0/1 + h_1w/1 + (1-h_1)h_2w/1 + (1-h_2)h_3w/1 + \dots$ , then the modulus of  $f_1(w)$  is bounded for  $|w| < 1$  if and only if the series

$$\sum [h_1h_2 \dots h_n/(1-h_1)(1-h_2) \dots (1-h_n)]$$

converges. Since by Theorem 3.2,  $h_n$  equals  $g_n$  or  $1-g_n$  according as  $n$  is even or odd, it will be seen that the latter series is the same as (3.11).

One may prove that  $(1+x)f(x)$  has its modulus bounded for  $\Re(x) > -\frac{1}{2}$  if and only if  $f(x)$  has a Stieltjes integral representation of the form  $\int_0^1 u d\phi(u)/(1+xu)$ ; and that the moduli of  $f(x)$  and of  $(1+x)f(x)$  are bounded in the unit circle, and in the half-plane  $\Re(x) > -\frac{1}{2}$ , respectively, if and only if the integral has the form  $\int_0^1 u(1-u) d\phi(u)/(1+xu)$ .

**4. An illustration.** In this section we offer an example to illustrate some of the results of the two preceding sections.

Let  $r$  be real,  $0 < r < 1$ , and consider the function

$$f(x) = \frac{1}{1} + \frac{r^2(1-r)x}{1} + \frac{r^3(1-r^2)x}{1} + \frac{r^4(1-r^3)x}{1} + \dots$$

We shall determine the moments generated by this function and the corresponding mass function  $\phi(u)$ .

Put  $F(x) = 1 + rxf(x)$ , so that  $F(x) = 1 + rx/1 + r^2(1-r)x/1 + r^3(1-r^2)x/1 + \dots$ . Then by Theorem 2.2 we have  $(1+x)/F(x) = 1 + (1-r)x/1 + r(1-r^2)x/1 + r^2(1-r^3)x/1 + \dots$  so that  $F(x)$  satisfies the functional equation

$$(4.1) \quad F(x) = 1 + \frac{rx}{1+rx} F(r^2x).$$

If we put  $f(x) = c_0 - c_1x + c_2x^2 - \dots$ , we obtain quite readily the values of the moments  $c_n$  from this functional relation:



$$(4.2) \quad c_0 = 1, \quad c_n = r^{2n}(1-r)(1-r^3) \cdots (1-r^{2n-1}), \quad n = 1, 2, 3, \dots$$

We obtain the following items of information about the corresponding mass function  $\phi(u)$ .

(a) Since  $\lim_{n \rightarrow \infty} r^n(1-r^{n-1}) = 0$ , it follows from a result of Stieltjes [8, p. 560] that  $f(x)$  is a meromorphic function of  $x$ . Consequently  $\phi(u)$  is a *step function* (with infinitely many discontinuities).

(b) Since  $\lim c_n = 0$ ,  $\phi(u)$  is continuous at  $u = 1$  by Theorem 3.5.

(c) From a result of Wall [6, p. 172] there exist numbers  $g_1, g_2, g_3, \dots$ ,  $0 < g_n < 1$ , such that  $g_1 = r^2(1-r)$ ,  $g_2(1-g_1) = r^3(1-r^2)$ ,  $g_3(1-g_2) = r^4(1-r^3)$ ,  $\dots$ . Then we find that the test-ratio for the series of Theorem 2.1 is

$$\frac{g_{2n-1}(1-g_{2n-2})}{g_{2n}(1-g_{2n-1})} = \frac{1-r^{2n-1}}{r(1-r^{2n})},$$

which has the limit  $(1/r) > 1$  as  $n \rightarrow \infty$ . Since the series then diverges, we conclude that  $\phi(u)$  is continuous at  $u = 0$ .

(d) In order to locate the discontinuities of  $\phi(u)$  we have to locate the poles of  $f(x)$ . From (4.1) we have the formal expansion

$$(4.3) \quad F(x) = 1 + \frac{rx}{1+r^2x} + \frac{r^4x^2}{(1+r^2x)(1+r^4x)} + \dots \\ + \frac{r^n}{(1+r^2x)(1+r^4x) \cdots (1+r^{2n}x)} + \dots$$

The series on the right converges for all  $x \neq -1/r^{2n}$ , ( $n = 1, 2, 3, \dots$ ), and is uniformly convergent in any bounded region from which the interiors of small circles about these points have been removed. Let  $S_n$  denote the sum of the first  $n$  terms of this series. Then, from (4.1), we have

$$F(x) = S_n + \frac{r^{n^2}x^n}{(1+r^2x)(1+r^4x) \cdots (1+r^{2n}x)} F(r^{2n}x), \quad n = 1, 2, 3, \dots$$

Now by Theorem 2.3  $|F(x)|$  is bounded for  $|x| \leq 1$ . Consequently  $\lim_{n \rightarrow \infty} S_n = F(x)$  for  $|x| \leq 1$ . It follows that (4.3) is a valid expansion of the function  $F(x)$ .

Since  $F(x) = 1 + rx f(x)$ , we have

$$f(x) = \sum_{n=1}^{\infty} \frac{r^{n(n-1)}x^{n-1}}{(1+r^2x)(1+r^4x) \cdots (1+r^{2n}x)},$$

for all  $x \neq -1/r^2, -1/r^4, \dots$ , and the latter points are the poles of  $f(x)$ .

(e) From (d) it follows that  $f(x)$  must have an expansion of the form

$$f(x) = \sum_{n=1}^{\infty} \frac{M_n}{1+r^{2n}x},$$

where  $M_n > 0$ , ( $n = 1, 2, 3, \dots$ ). Thus, the function  $\phi(u)$  will be completely determined at all points of continuity (except for an additive constant) if we know the values of the numbers  $M_n$ . By (4.1) we have  $(1+r^2x)f(x) = 1+r^3xf(r^2x)$ , and therefore

$$\begin{aligned} M_1 &= 1 - rf(-1) = 1 - r - r^3(1-r) - r^5(1-r)(1-r^3) - \dots \\ &= \prod_{n=1}^{\infty} (1 - r^{2n-1}). \end{aligned}$$

Then, by this same relation,

$$(4.4) \quad M_n = \frac{r}{1 - r^{2n-2}} M_{n-1}, \quad n = 2, 3, 4, \dots$$

This determines  $\phi(u)$  in accordance with the following table of values.

Value of $\phi(u)$	Value of $u$
1	$r^2 \leq u \leq 1$
$1 - M_1$	$r^4 \leq u < r^2$
$1 - \left(1 + \frac{r}{1 - r^2}\right) M_1$	$r^6 \leq u < r^4$
$1 - \left(1 + \frac{r}{1 - r^2} + \frac{r^2}{(1 - r^2)(1 - r^4)}\right) M_1$	$r^8 \leq u < r^6$
...	...
0	$u = 0$

5. **Hypergeometric summability.** The continued fraction of Gauss [11, p. 348] generates an interesting totally monotone sequence when the parameters are properly restricted. Thus, if we have given the special hypergeometric series

$$F(\alpha, 1, \gamma, x) = 1 + \frac{\alpha}{\gamma} x + \frac{\alpha(\alpha+1)}{\gamma(\gamma+1)} x^2 + \dots,$$

we can obtain the representation

$$F(\alpha, 1, \gamma, -x) = 1/1 + g_1x/1 + (1 - g_1)g_2x/1 + (1 - g_2)g_3x/1 + \dots,$$

where

$$g_{2n} = \frac{n}{\gamma + 2n - 1}, \quad g_{2n-1} = \frac{\alpha + n - 1}{\gamma + 2n - 2}, \quad n = 1, 2, 3, \dots$$

The  $g_n$ 's will be real and will lie between 0 and 1 if and only if  $\alpha$  and  $\gamma$  are real,  $\gamma > \alpha > 0$ . The moments are then

$$(5.1) \quad c_0 = 1, \quad c_n = \frac{\alpha(\alpha+1)(\alpha+2) \cdots (\alpha+n-1)}{\gamma(\gamma+1)(\gamma+2) \cdots (\gamma+n-1)}, \quad n = 1, 2, 3, \dots$$

Since the hypergeometric series converges for  $x=1$ ,  $\gamma > \alpha$ , it follows that  $\lim c_n = 0$ , and hence that the mass function  $\phi(u)$  is continuous at  $u=1$ . Moreover, it is easy to show that the series (2.3) diverges, and hence that  $\phi(u)$  is continuous at  $u=0$ . Accordingly, the moment sequence generated by this continued fraction is regular when  $\alpha, \gamma$  are real and  $\gamma > \alpha > 0$ .

We can determine  $\phi(u)$  from the familiar Eulerian integral of the first kind. In fact, we have

$$F(\alpha, 1, \gamma, -x) = \int_0^1 \frac{d\phi(u)}{1+xu},$$

where

$$\phi(u) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)} \int_0^u t^{\alpha-1}(1-t)^{\gamma-\alpha-1} dt.$$

The sequence (5.1) is a special case of the sequence of coefficients in the general hypergeometric series:

$$(5.2) \quad c_0 = 1, \\ c_n = \frac{\alpha(\alpha+1)(\alpha+2) \cdots (\alpha+n-1)\beta(\beta+1)(\beta+2) \cdots (\beta+n-1)}{\gamma(\gamma+1)(\gamma+2) \cdots (\gamma+n-1) \cdot 1 \cdot 2 \cdot 3 \cdots n}, \\ n = 1, 2, 3, \dots$$

Accordingly, it is convenient to designate the method of summability defined by (5.2), when the sequence is regular, as *hypergeometric summability*. In this connection we use the symbol  $(H, \alpha, \beta, \gamma)$ , where in particular the sequence (5.1) defines summability  $(H, \alpha, 1, \gamma)$ .

By means of (1.4) the general term of the Hausdorff matrix associated with summability  $(H, \alpha, 1, \gamma)$  is readily found to be

$$(5.3) \quad \frac{C_{\alpha+n-1,n} C_{m-n+\gamma-\alpha-1,m-n}}{C_{\gamma+m-1,m}}, \quad \gamma > \alpha > 0.$$

Next we display some of the inclusion and equivalence relationships between the hypergeometric methods. Some of the symbolism is not readily intelligible and will be explained presently.

- (5.4) (i)  $(H, \alpha, 1, \gamma) = (H, 1, \alpha, \gamma)$ ,  $\gamma > \alpha$ ;  
 (ii)  $(H, \alpha, \gamma, \gamma) = (H, \alpha, 1, 1)$ ,  $0 < \alpha < 1$ ;  $\gamma > 0$ ;  
 (iii)  $(H, 1, 1, \gamma+1) = (C, \gamma)$ ,  $\gamma > 0$ ;  
 (iv)  $(H, \alpha, 1, \gamma+1) \subset (C, \gamma)$ ,  $\alpha > 1$ ;  $\gamma > 0$ ;  $(H, 1, 1, \alpha)$ ;  
 (v)  $(H, \alpha, 1, \gamma+1) \supset (C, \gamma)$ ,  $0 < \alpha < 1$ ;  $\gamma > 0$ ;  $(H, \alpha, 1, 1)$ ;

- (vi)  $(H, \alpha, 1, \alpha+1) \subset (C, \alpha), \alpha > 1;$
- (vii)  $(H, \alpha, 1, \alpha+1) \supset (C, \alpha), 0 < \alpha < 1;$
- (viii)  $(H, \alpha, 1, \gamma_1) \subset (H, \alpha, 1, \gamma_2), \gamma_2 > \gamma_1 > \alpha > 0; (H, \gamma_1, 1, \gamma_2);$
- (ix)  $(H, \alpha_2, 1, \gamma) \subset (H, \alpha_1, 1, \gamma), \alpha_2 > \alpha_1 > 0; \gamma > 0; (H, \alpha_1, 1, \alpha_2);$
- (x)  $(C, \alpha_1) \subset (C, \alpha_2), \alpha_2 > \alpha_1 > -1; (H, \alpha_1 + 1, 1, \alpha_2 + 1);$
- (xi)  $(H, \alpha, 1, \alpha+1) \approx (H, \beta, 1, \beta+1); \alpha, \beta > 0;$
- (xii)  $(C, 1) \approx (H, \alpha, 1, \alpha+1), \alpha > 0;$
- (xiii)  $(H, \alpha_1, 1, \alpha_1 + \beta) \approx (H, \alpha_2, 1, \alpha_2 + \beta); \alpha_1, \alpha_2, \beta > 0; |\alpha_1 - \alpha_2| = 1, 2, 3, \dots;$
- (xiv)  $(H, \alpha, k_1 + 1, \gamma) \supset (H, \alpha, k_2 + 1, \gamma); k_1, k_2 = 0, 1, 2, \dots; k_2 > k_1; \gamma > \alpha + \kappa_2 > 1; (H, k_1 + 1, 1, k_2 + 1);$
- (xv)  $(H, \alpha, 1, \gamma) \supset (H, \alpha, k + 1, \gamma), k = 1, 2, 3, \dots; \gamma > \alpha + \kappa > 1; (H, 1, 1, k + 1).$

It should be mentioned that the matrix defined by (5.3) can be found in the literature on summability, but not in the same form. It was used by Cesàro in his celebrated theorem on the Cauchy product of two Cesàro summable series [12, p. 489], by Knopp in his proof of the equivalence of the  $(C, k)$  and  $(H, k)$  methods for positive integral values of  $k$  [12, p. 481], and by Hausdorff [3] in his proof of the same equivalence theorem. However, it has never before been associated with the sequence of coefficients in the hypergeometric series, and most of the relationships in (5.4) are new.

The identities (i) and (ii), and other related ones, obtain in virtue of the symmetric form of the sequence (5.2). The identity (iii) merely exhibits Cesàro summability as a special case of hypergeometric summability.

The remaining relationships are proved with the aid of the following theorems of Hausdorff [4].

**THEOREM 5.1.** *Necessary and sufficient conditions that a matrix*

$$\mathfrak{A} = \mathfrak{D}(\delta_{mn}c_m)\mathfrak{D} = (C_{m,n}\Delta^{m-n}c_n)$$

*be regular are*

- (i)  $c_0 = 1,$
- (ii)  $\sum_{n=0}^m C_{m,n} |\Delta^{m-n}c_n| \leq M, \quad m = 0, 1, 2, \dots,$
- $M$  independent of  $m,$
- (iii)  $\lim_{m \rightarrow \infty} C_{m,n}\Delta^{m-n}c_n = 0, \quad n = 0, 1, 2, \dots.$

**THEOREM 5.2.** *Let  $\mathfrak{A} = \mathfrak{D}(\delta_{mn}c_m^A)\mathfrak{D}$  and  $\mathfrak{B} = \mathfrak{D}(\delta_{mn}c_m^B)\mathfrak{D}$  be regular matrices, and let  $\mathfrak{B}^{-1}$  exist. Then, a necessary and sufficient condition that  $\mathfrak{A} \supset \mathfrak{B}$  is that the matrix  $\mathfrak{D}(\delta_{mn}c_m^A/c_m^B)\mathfrak{D}$  be regular.*

**THEOREM 5.3.** *Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be regular matrices, and let  $\mathfrak{A}^{-1}$  and  $\mathfrak{B}^{-1}$  exist. Then, necessary and sufficient conditions that  $\mathfrak{A} \approx \mathfrak{B}$  are that the matrices*

$$\mathfrak{D}(\delta_{mn} \frac{A}{c_m}/\frac{B}{c_m})\mathfrak{D}, \quad \mathfrak{D}(\delta_{mn} \frac{B}{c_m}/\frac{A}{c_m})\mathfrak{D}$$

*be regular.*

The relation (iv) is established with the aid of Theorem 5.2. Let  $\{c'_n\}$  and  $\{c''_n\}$  be the moment sequences associated respectively with the  $(C, \gamma)$  and  $(H, \alpha, 1, \gamma+1)$  methods of summation. Then

$$\begin{aligned} \frac{c'_n}{c''_n} &= \frac{1 \cdot 2 \cdots n}{(\gamma+1)(\gamma+2) \cdots (\gamma+n)} \div \frac{\alpha(\alpha+1) \cdots (\alpha+n-1)}{(\gamma+1)(\gamma+2) \cdots (\gamma+n)} \\ &= \frac{1 \cdot 2 \cdots n}{\alpha(\alpha+1) \cdots (\alpha+n-1)}. \end{aligned}$$

However, this defines the regular moment sequence associated with summability  $(H, 1, 1, \alpha)$ ,  $(\alpha > 1)$ . This completes the proof, and explains the appendage to statement (iv). The relations (v), (viii), (ix), and (x) are established with the same technique.

The statement (v) is of particular interest due to the scarcity of methods of summation of the type (1.1) which include  $(C, \gamma)$  summability. Indeed, we know of only two commonly known methods of the type (1.1) which have this property, the method of de la Vallée Poussin<sup>(4)</sup> and a method of M. Riesz<sup>(5)</sup>. Hille and Tamarkin [15] give an interesting set of necessary and sufficient conditions in order that a Hausdorff method shall include  $(C, \gamma)$ . They are stated in detail for the case when  $\gamma$  is an integer, and the conditions are easily handled.

It is convenient at this time to define Riesz means of the first order and Nörlund means. We write

$$(5.5) \quad \frac{p_0 s_0 + p_1 s_1 + \cdots + p_n s_n}{P_n},$$

$$(5.6) \quad \frac{p_n s_0 + p_{n-1} s_1 + \cdots + p_0 s_n}{P_n},$$

where  $P_n = p_0 + p_1 + \cdots + p_n$ ,  $(p_n \geq 0)$ , and  $\sum p_n$  in (5.5) is always divergent. Since means of the type (5.5) were used in the early development of Riesz typical means, they are called Riesz means and are designated by  $(R, p_n)$ .

<sup>(4)</sup> That the method of de la Vallée Poussin includes  $(C, \gamma)$  summability was proved independently and virtually simultaneously by T. H. Gronwall [13, p. 1664], and C. N. Moore [14, p. 1774].

<sup>(5)</sup> It has been proved that the Riesz logarithmic mean of order  $\gamma$  provides a method of summation which includes  $(C, \gamma)$  summability [16].

Means of the type (5.6) are called Nörlund means and are designated by  $(N, p_n)$ .

The statements (vi) and (vii) are of course special cases of (iv) and (v) respectively. It is of interest that the transform (1.1) associated with  $(H, \alpha, 1, \alpha+1)$ , a method of the Riesz type, contains the coefficients in the transform associated with  $(C, \alpha)$ , a method of the Nörlund type, written in the reverse order.

The statement (x) is a classical result in the domain of Cesàro summability. The proof is particularly easy to understand with the aid of the hypergeometric notation.

In order to prove (xi), (xii), and (xiii) we need the lemma which follows.

**LEMMA 5.1.** *Let the sequences  $\{c_n^i\}$ ,  $(i=1, 2, 3, \dots, k)$ , be regular sequences. Then, if  $\sum_{i=1}^k d_i = 1$ , the sequence  $\{c_n\}$ , where*

$$c_n = \sum_{i=1}^k d_i c_n^i, \quad n = 0, 1, 2, \dots,$$

*is also regular.*

The condition imposed on the  $c_i$ 's insures that condition (i) of Theorem 5.1 be fulfilled. The other conditions of the theorem will be fulfilled since the difference operation is linear.

To start the proof of (xi), let  $\{c_n^\alpha\}$  and  $\{c_n^\beta\}$  be the moment sequences associated respectively with the methods  $(H, \alpha, 1, \alpha+1)$  and  $(H, \beta, 1, \beta+1)$ . Then

$$\frac{c_n^\alpha}{c_n^\beta} = \frac{\alpha(\beta+n)}{\beta(\alpha+n)} = \frac{\alpha}{\beta} + \frac{\beta-\alpha}{\beta} \cdot \frac{\alpha}{\alpha+n}, \quad \alpha, \beta > 0.$$

Let  $\{c_n'\} = (1, 1, 1, \dots)$ . Then

$$c_n^\alpha / c_n^\beta = d_1 c_n' + d_2 c_n^\beta,$$

where  $d_1 = \beta/\alpha$ ,  $d_2 = (\alpha-\beta)/\alpha$ ,  $d_1 + d_2 = 1$ . Then, by Lemma 5.1, the sequence  $\{c_n^\alpha / c_n^\beta\}$  is regular. Using Theorem 5.2 we now have  $(H, \alpha, 1, \alpha+1) \supset (H, \beta, 1, \beta+1)$ . By repetition of this argument we can also show that the sequence  $\{c_n^\beta / c_n^\alpha\}$  is regular. Thus, we have  $(H, \alpha, 1, \alpha+1) \subset (H, \beta, 1, \beta+1)$ . This proof is due to Hausdorff [3].

The relation (xii) is an interesting special case of (xi). We observe that summability  $(H, \alpha, 1, \alpha+1)$ ,  $(\alpha > 0)$ , is essentially summability  $(R, n^{\alpha-1})$ ,  $(\alpha > 0)$ , and that the relationship  $(C, 1) \approx (R, n^{\alpha-1})$ ,  $(\alpha > 0)$ , can be proved in a completely different fashion [17].

Knopp [12, p. 481] has proved (xiii) and (viii) by very laborious methods for integral values of the parameters. These statements afford the entire basis for his Cesàro-Hölder equivalence proof. To establish (xiii) we assume that



$\alpha_2 > \alpha_1$  and first prove that  $(H, \alpha_1, 1, \alpha_1 + \beta) \approx (H, \alpha_1 + 1, 1, \alpha_1 + 1 + \beta)$ . Let  $\{c_n'\}$  and  $\{c_n''\}$  be the moment sequences associated respectively with these two methods. Then

$$\begin{aligned} \frac{c_n'}{c_n''} &= \frac{\alpha_1}{\alpha_1 + n} \cdot \frac{\alpha_1 + \beta + 1 + n}{\alpha_1 + \beta + 1} \\ &= \frac{\alpha_1}{\alpha_1 + \beta + 1} + \frac{\beta + 1}{\alpha_1 + \beta + 1} \cdot \frac{\alpha_1}{\alpha_1 + n}, \quad n = 0, 1, 2, \dots, \end{aligned}$$

and is consequently a linear combination of regular sequences, where the constants of combination add up to unity. Likewise, as in the proof of (xi),  $\{c_n''/c_n'\}$  is a regular sequence. We use Theorem 5.3 to complete the first stage of the proof. Next we prove in the same fashion that  $(H, \alpha_1 + 1, 1, \alpha_1 + 1 + \beta) \approx (H, \alpha_1 + 2, 1, \alpha_1 + 2 + \beta)$ . Clearly, it can readily be established by induction that  $(H, \alpha_1, 1, \alpha_1 + \beta) \approx (H, \alpha_2, 1, \alpha_2 + \beta)$ , provided that  $\alpha_1$  and  $\alpha_2$  differ by an integer.

Up to the present time we have considered hypergeometric summability  $(H, \alpha, \beta, \gamma)$ , only for the case  $\beta = 1, \gamma > \alpha$ . It is possible to find non-trivial regular hypergeometric methods at least for positive integral values of  $\beta$  by use of a known relationship [18, p. 233] among the hypergeometric functions. If we write

$$F(\alpha, \beta, \gamma, x) = 1 + \frac{\alpha\beta}{\gamma \cdot 1} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1) \cdot 1 \cdot 2} x^2 + \dots,$$

the series always converges and represents the function for  $|x| < 1; \alpha, \beta, \gamma > 0$ . Now, we consider the identity

$$(5.7) \quad (\beta - \alpha)F(\alpha, \beta, \gamma, x) + \alpha F(\alpha + 1, \beta, \gamma, x) - \beta F(\alpha, \beta + 1, \gamma, x) = 0, \quad |x| < 1.$$

Set  $\beta = 1$  in (5.7) and write

$$F(\alpha, 2, \gamma, x) = (1 - \alpha)F(\alpha, 1, \gamma, x) + \alpha F(\alpha + 1, 1, \gamma, x), \quad |x| < 1.$$

Equating coefficients of  $x^n$  in this identity we obtain the relation

$$(5.8) \quad \frac{C_{\alpha+n-1,n}(n+1)}{C_{\gamma+n-1,n}} = \frac{(1-\alpha)C_{\alpha+n-1,n}}{C_{\gamma+n-1,n}} + \frac{\alpha C_{\alpha+n,n}}{C_{\gamma+n-1,n}}, \quad n = 0, 1, 2, \dots$$

Now, the expression on the right is a linear combination of regular sequences, provided that  $\gamma > \alpha + 1; \alpha > 0$ , where the constants of combination add up to unity. Consequently, the left member of (5.8) defines a regular method of summability,  $(H, \alpha, 2, \gamma), \gamma > \alpha + 1; \alpha > 0$ . If we now use (5.7) as a recursion relation, we can define summability  $(H, \alpha, 3, \gamma)$  in terms of summability  $(H, \alpha, 2, \gamma)$ . Continuing this process, we define the regular methods

$(H, \alpha, k+1, \gamma)$ ,  $(k=0, 1, 2, \dots)$ ;  $\gamma > \alpha + \kappa$ ;  $\alpha > 0$ , with the associated moment sequence

$$\frac{C_{\alpha+n-1,n} C_{k+n,n}}{C_{\gamma+n-1,n}}, \quad n = 0, 1, 2, \dots$$

It is now easy to prove the statement (xiv) using methods already established. Statement (xv), a special case of (xiv), clearly indicates the weakness of these newly defined methods.

6. **Special methods in the difference matrix.** In this section we propose to replace the base sequence  $\{c_n\}$  in the difference matrix  $\Delta = (\Delta^m c_n)$  by known regular sequences whose corresponding mass functions are continuous at  $u=1$ , and then discuss the resulting methods of summation.

*Rows in the difference matrix.* If in the matrix  $\Delta$  the regular sequence associated with summability  $(H, \alpha, 1, \gamma)$  is taken as the base sequence, then the  $(k+1)$ -st row defines summability  $(H, \alpha, 1, \gamma+k)$ . It is understood of course, that we normalize each row sequence by dividing each member of the sequence by its first term. Using (5.4, viii) we see that the efficiency of the new methods increases with the depth of the row in the matrix.

If we start with the Euler-Knopp sequence,  $\{\theta^n\}$ ,  $0 < \theta < 1$ , which defines summability  $(E, \theta)$ , as the base sequence, no change occurs as a result of repeated differencing.

E. Hille has given us an interesting example to prove that repeated differencing of the base sequence does not always improve or leave unchanged the efficiency of a Hausdorff method corresponding to a monotone non-decreasing mass function  $\phi(u)$ . In this connection he utilizes the integral

$$c(z) = \int_0^1 u^z d\phi(u), \quad \Re(z) \geq 0,$$

which is called the *moment function* of the associated Hausdorff method. The function  $c(z)$  is holomorphic when  $\Re(z) > 0$ , and it is continuous in  $\Re(z) \geq 0$  [15]. To show that a Hausdorff method  $[H, \phi_1(u)]$  includes  $[H, \phi_2(u)]$  we must establish the existence of a moment function  $c(z)$  such that

$$c_1(z) = c(z)c_2(z).$$

Now, let  $\phi(u)$  be a step function with two discontinuities so that

$$c_n = \alpha a^n + \beta b^n, \quad \alpha + \beta = 1, \quad \beta < \alpha; \quad 0 < a < b < 1.$$

Thus,  $\phi(u)$  is a monotone non-decreasing function. The moment function  $c(z) = \alpha a^z + \beta b^z$  has a set of equidistant zeros on a vertical line in the right half-plane. The moment function corresponding to the normalized sequence of the  $m$ th row of the difference matrix is  $c_m(z) = \Delta^m c(z) / \Delta^m c(0)$ . Now, if the Hausdorff method of summation defined by the sequence  $\{c_m(n)\}$  is to in-

clude or be equivalent to the method defined by  $\{c(n)\}$ , then the quotient  $c_m(z)/c(z)$  has to be a moment function. In particular, it must be holomorphic for  $\Re(z) > 0$ . However,  $c(z)$  vanishes for

$$z = \left[ \log \frac{\alpha}{\beta} + (2k+1)\pi i \right] \left( \log \frac{b}{a} \right)^{-1}, \quad k = 0, \pm 1, \pm 2, \dots,$$

and  $c_m(z)$  vanishes for

$$z = \left\{ \log \left[ \frac{\alpha}{\beta} \left( \frac{1-a}{1-b} \right)^m \right] + (2k+1)\pi i \right\} \left( \log \frac{b}{a} \right)^{-1},$$

$$k = 0, \pm 1, \pm 2, \dots$$

Since  $(1-a)/(1-b) > 1$ , the zeros of  $c(z)$  and  $c_m(z)$  are completely distinct, and their quotient is not holomorphic in the right half-plane. This establishes the case in point.

Hille's example raises the problem of determining conditions on the mass function  $\phi(u)$  in order that repeated differencing of the base sequence will yield new methods of unchanging or steadily increasing efficiency.

*Columns in the difference matrix.* If we use the regular moment sequence associated with summability  $(C, \alpha)$  as the base sequence in  $\Delta$ , the  $(k+1)$ -st column defines summability  $(H, \alpha, 1, \alpha+k+1)$ . These methods increase in efficiency with increasing  $k$ . Moreover, from (5.4, xiii), we see that if we vary  $\alpha$  by any integral amount  $\pm p$ , ( $p=1, 2, 3, \dots$ ), such that  $\alpha \pm p > 0$ , the efficiency of the new methods, corresponding to a particular  $k$ , remains unchanged. If we start with the sequence associated with  $(H, \alpha, 1, \gamma)$  as the base sequence, we get summability  $(H, \gamma-\alpha, 1, k+\gamma)$  in the  $(k+1)$ -st column. Again, the efficiency of the methods increases as we traverse the matrix to the right by columns.

Starting with the Euler sequence as a base sequence we get expected results. We obtain the sequence defined by  $(1-\theta)^n$ ,  $0 < \theta < 1$ , in every column. Thus there is no increase in efficiency with an increasing column index.

*Diagonal files in the difference matrix.* It has already been established that the diagonal files of  $\Delta$ , as well as the rows and columns, yield regular moment sequences provided the mass function of the base sequence satisfies appropriate continuity requirements. However, starting with familiar base sequences, the diagonal files yield regular moment sequences of a new type. It is of interest to recall that the mass function associated with these new sequences is constant for  $\frac{1}{2} \leq u \leq 1$ . As a result of the discussion in the next section it will be established that all of the diagonal files define methods of summation which include  $(E, \frac{1}{2})$ .

If we start with the sequence associated with summability  $(C, \alpha)$  as the base sequence, the *principal diagonal* then yields the regular moment sequence

$$(6.1) \quad \frac{\alpha}{(n+1)C_{2n+\alpha, n+1}}, \quad n = 0, 1, 2, \dots; \alpha > 0.$$

If we start with summability  $(C, \beta)$  we obtain in the  $\nu$ th upper diagonal the regular moment sequence

$$(6.2) \quad \frac{\beta C_{\nu+\beta, \beta}}{(n+\nu+1)C_{2n+\nu+\beta, n+\nu+1}}, \quad n = 0, 1, 2, \dots; \beta > 0.$$

If we start with summability  $(C, \gamma)$  the  $\mu$ th lower diagonal gives the sequence

$$(6.3) \quad \frac{\mu + \gamma}{(n+1)C_{2n+\mu+\gamma, n+1}}, \quad n = 0, 1, 2, \dots; \gamma > 0.$$

Observe that  $\mu$  and  $\nu$  vary through positive integral values only. If  $\alpha, \beta, \gamma$  also vary through only positive integral values, all of the methods of summation defined by the sequences (6.1), (6.2), and (6.3) are equivalent. Indeed, these methods still remain equivalent to each other if  $\alpha, \beta, \gamma$  vary in any fashion so as to differ from each other only by integral amounts. These statements may be proved by using the same technique as employed in proving (5.4, xiii). To illustrate the procedure used, let  $\{c'_n\}$  and  $\{c''_n\}$  be respectively the sequence (6.3) for fixed  $\mu$  and  $\gamma$ , and the sequence obtained from (6.3) by replacing  $\gamma$  by  $\gamma+1$ . We propose to show that the methods of summation defined by these sequences are equivalent. We form the quotient

$$\frac{c'_n}{c''_n} = \frac{\mu + \gamma}{n + \mu + \gamma} \cdot \frac{n + \frac{1}{2}(\mu + \gamma + 1)}{\frac{1}{2}(\mu + \gamma + 1)}.$$

In the light of past experience, we see that both  $\{c'_n/c''_n\}$  and  $\{c''_n/c'_n\}$  are regular sequences. Thus, the methods of summation in question are equivalent.

Starting with the Knopp-Euler sequence as a base sequence, we obtain the regular sequence  $\{(1-\theta)^n \theta^n\}$ ,  $0 < \theta < 1$ , in all the diagonals. This is clearly another Knopp-Euler sequence.

*Symmetry in the difference matrix.* It is of some interest to find regular sequences which give rise to symmetry about the principal diagonal of the difference matrix. For the hypergeometric method  $(H, \alpha, 1, \gamma)$  we have

$$\Delta^m c_n = \frac{C_{\alpha+n-1, n} C_{\gamma-\alpha+m-1, m}}{C_{\gamma+m+n-1, m+n}}.$$

Thus, the method  $(H, 1, 1, \gamma) = (C, \gamma-1, (\gamma > 1))$ , and the method  $(H, \frac{1}{2}, 1, 1)$  have this property. Moreover, the Knopp-Euler method  $(E, \frac{1}{2})$  also has this property.

We first noticed the symmetry of the method  $(H, \frac{1}{2}, 1, 1)$  while considering the periodic continued fraction

$$\frac{1}{1 + \frac{rx}{1 + \frac{r(1-r)x}{1 + \frac{r(1-r)x}{1 + \dots}}}}, \quad 0 > r > 1.$$

The function represented by this continued fraction is

$$f(x) = \frac{(1-2r) - [1 + 4r(1-r)x]^{1/2}}{-2r(1+x)} = c_0 - c_1x + c_2x^2 - \dots.$$

Then

$$\Delta c_0 - \Delta c_1x + \Delta c_2x^2 - \dots = \frac{1 - [1 + 4r(1-r)x]^{1/2}}{-2rx}.$$

Hence

$$(6.4) \quad \Delta c_n = \frac{(2n)!}{n!(n+1)!} r^n (1-r)^{n+1}, \quad n = 0, 1, 2, \dots$$

The moment sequence (6.4) for  $r = \frac{1}{2}$ , when normalized, is the moment sequence for summability  $(H, \frac{1}{2}, 1, 1)$ .

It is easy to prove by means of Theorem 3.2 that the difference matrix is symmetric about the principal diagonal if and only if the function  $f(x)$  which generates the base sequence has a continued fraction of the form given in (1.5, iii) in which  $g_{2n-1} = \frac{1}{2}$ ,  $(n = 1, 2, 3, \dots)$ .

**7. Analytic continuation.** This section is devoted to some remarks concerning the effectiveness of the Hausdorff methods in the analytic continuation of a power series  $\sum a_n z^n$  outside of its conventional circle of convergence. The Hausdorff methods of particular interest in this respect are those for which the mass function  $\phi(u)$  is a monotone non-decreasing function which is constant in the neighborhood of  $u = 1$ .

The Hausdorff transform of the sequence  $\{s_n\}$  is given by

$$(7.1) \quad \sigma_m = \sum_{n=0}^m C_{m,n} \Delta^{m-n} c_n \cdot s_n = \int_0^1 \sum_{n=0}^m C_{m,n} u^n (1-u)^{m-n} s_n d\phi(u).$$

If  $\phi(u)$  is a monotone non-decreasing function satisfying the regularity conditions (5.5), which is constant for  $\delta < u \leq 1$ , we shall designate (7.1) the  $\delta$ -transform of the sequence  $\{s_n\}$ .

Now, from a result of Hille and Tamarkin [15], a necessary and sufficient condition that  $[H, \phi(u)] \supset (E, \delta)$  is that  $\phi(u) = 1$ ,  $\delta \leq u \leq 1$ . Thus, the  $\delta$ -method has at least the same efficiency as the Euler-Knopp method in the problem of analytic continuation. It will be of interest to recall the nature of this region of convergence. Corresponding to each singularity  $\zeta$  of the power

series, draw the circle whose equation is

$$\left| \frac{z}{\xi} + \frac{1-\delta}{\delta} \right| = \frac{1}{\delta}.$$

These circles are tangent to the sides of the Borel polygon of summability at the points of singularity. The figure thus constructed is called the *curvilinear polygon*  $B_\delta$ .

Knopp [19] has established  $B_\delta$  as the region of convergence for the method  $(E, \delta)$  for the case that  $\delta = 2^{-p}$ , ( $p = 1, 2, 3, \dots$ ). Using the methods of Knopp, we can readily prove that the  $\delta$ -transform of  $f(z) = \sum a_n z^n$  converges to  $f(z)$  at a point  $z$  inside of  $B_\delta$ , but we have been unable to establish divergence outside  $B_\delta$ . However, we shall offer some evidence in support of the following *conjecture*: a necessary and sufficient condition that a Hausdorff method of summation, corresponding to a monotone non-decreasing  $\phi(u)$  satisfying the regularity conditions (1.3), shall sum a power series outside of its circle of convergence is that  $\phi(u)$  be constant in the neighborhood of  $u=1$ .

In support of our conjecture we shall first prove that if  $\phi(u)$  is not constant in the neighborhood of  $u=1$  then the  $\delta$ -transform of the geometric series  $\sum z^n$  diverges for  $z = -1 - \epsilon$ ,  $\epsilon > 0$ . We have for this case

$$\begin{aligned} \sigma_m &= \frac{1}{1-z} - \frac{z}{1-z} \int_0^1 \sum_{n=0}^m C_{m,n}(uz)^n (1-u)^{m-n} d\phi(u) \\ &= \frac{1}{1-z} - \frac{z}{1-z} \int_0^1 (uz + 1 - u)^m d\phi(u). \end{aligned}$$

If  $z = -1 - \epsilon$ , then

$$\sigma_m = \frac{1}{2+\epsilon} + \frac{1+\epsilon}{2+\epsilon} \int_0^1 (1-2u-\epsilon u)^m d\phi(u),$$

and

$$\left| \sigma_m - \frac{1}{2+\epsilon} \right| = \frac{1+\epsilon}{2+\epsilon} \left| \int_0^1 (1-2u-\epsilon u)^m d\phi(u) \right|.$$

Let  $m$  be even. Then, for  $\epsilon_1 > 0$ ,  $\epsilon_1 < \epsilon$ ,  $\eta = (2+\epsilon_1)/(2+\epsilon)$ , we have

$$\left| \sigma_m - \frac{1}{2+\epsilon} \right| > \frac{1}{2} \int_\eta^1 (1+\epsilon_1)^m d\phi(u) = (1+\epsilon_1)^m [\phi(1) - \phi(\eta)].$$

Hence,  $|\sigma_m| \rightarrow \infty$  as  $m \rightarrow \infty$ .

Next, we shall sum the series  $\sum z^n$  with a special  $\delta$ -method. Incidentally, the regular sequences (6.1), (6.2), and (6.3) define  $\delta$ -methods. Other examples are readily constructed. As a case in point, let  $\phi(u) = u/\delta$ ,  $0 \leq u \leq \delta$ ,



$0 < \delta < 1$ ;  $\phi(u) = 1$ ,  $\delta \leq u \leq 1$ . Let us test the corresponding  $\delta$ -transform on the series  $\sum z^n$ . We have

$$c_n = \frac{1}{\delta} \int_0^\delta u^n du = \frac{\delta^n}{n+1},$$

and

$$\Delta^{m-n} c_n = \frac{1}{\delta} \int_0^\delta u^n (1-u)^{m-n} du.$$

Then, the  $\delta$ -transform of the series  $\sum z^n$  is

$$\begin{aligned} \sum_{n=0}^m C_{m,n} \Delta^{m-n} c_n s_n &= \frac{1}{1-z} - \frac{z}{\delta(1-z)} \int_0^\delta \sum_{n=0}^m C_{m,n} (uz)^n (1-u)^{m-n} du \\ &= \frac{1}{1-z} + \frac{z(\delta z + 1 - \delta)^{m+1}}{\delta(1-z)^2(m+1)} - \frac{z}{\delta(1-z)^2(m+1)}. \end{aligned}$$

Clearly the transform converges to  $1/(1-z)$  as  $m \rightarrow \infty$  whenever

$$(7.2) \quad \left| z + \frac{1-\delta}{\delta} \right| < \frac{1}{\delta},$$

and diverges to  $+\infty$  whenever  $|z + (1-\delta)/\delta| > 1/\delta$ . Evidently, the region  $B_\delta$  is the largest region of convergence for the special case under consideration. Finally, the fact that the inequality (7.2) becomes  $|z| < 1$  as  $\delta \rightarrow 1$  is consistent with our conjecture.

#### REFERENCES

1. L. L. Silverman, *On the Definition of the Sum of a Divergent Series*, University of Missouri Studies, Mathematics Series, vol. 1, no. 1, 1913.
2. O. Toeplitz, *Über allgemeine lineare Mittelbildungen*, Prace Matematyczno-fizyczne, vol. 22 (1911), pp. 131-119.
3. F. Hausdorff, *Summationsmethoden und Momentfolgen*, I and II, Mathematische Zeitschrift, vol. 9 (1921), pp. 74-109, 280-299.
4. H. L. Garabedian, *Hausdorff matrices*, American Mathematical Monthly, vol. 46 (1939), pp. 390-410.
5. F. Hausdorff, *Über das Momentenproblem für ein endliches Intervall*, Mathematische Zeitschrift, vol. 16 (1923), pp. 220-248.
6. H. S. Wall, *Continued fractions and totally monotone sequences*, these Transactions, vol. 48 (1940), pp. 165-184.
7. T. J. Stieltjes, *Recherches sur les fractions continues*, Annales de l'Université de Toulouse, vol. 8, J (1894), pp. 1-122; vol. 9, A (1895), pp. 1-47.
8. T. J. Stieltjes, *Oeuvres*, vol. 2, Société Mathématique d'Amsterdam, Groningen, P. Noordhoff, 1918.
9. W. T. Scott and H. S. Wall, *A convergence theorem for continued fractions*, these Transactions, vol. 47 (1940), pp. 155-172.
10. I. J. Schoenberg, *Über die asymptotische Verteilung reeller Zahlen mod 1*, Mathematische Zeitschrift, vol. 28 (1928), pp. 171-199.
11. O. Perron, *Die Lehre von den Kettenbrüchen*, Leipzig and Berlin, Teubner, 1913.

12. K. Knopp, *Theorie und Anwendung der unendlichen Reihen*, Berlin, Springer, 1922, 2d edition, 1924.
13. T. H. Gronwall, *Comptes Rendus de l'Académie des Sciences*, Paris, vol. 158 (1914), p. 1664.
14. C. N. Moore, *ibid.*, vol. 158 (1914), p. 1774.
15. E. Hille and J. D. Tamarkin, *Questions of relative inclusion in the domain of Hausdorff means*, *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 573-577.
16. G. H. Hardy and M. Riesz, *The General Theory of Dirichlet's Series*, Cambridge Mathematical Tracts, no. 18, 1915.
17. H. L. Garabedian and W. C. Randels, *Theorems on Riesz means*, *Duke Mathematical Journal*, vol. 4 (1938), pp. 529-533.
18. J. Pierpont, *The Theory of Functions of a Complex Variable*, vol. 2, New York, Ginn, 1912.
19. K. Knopp, *Über das Eulersche Summierungsverfahren I*, *Mathematische Zeitschrift*, vol. 15 (1922), pp. 226-253; II, *ibid.*, vol. 18 (1923), pp. 125-156.

NORTHWESTERN UNIVERSITY,  
EVANSTON, ILL.

# POLYADIC GROUPS

BY

EMIL L. POST

## TABLE OF CONTENTS

SECTION	PAGE
Introduction . . . . .	209
I. GENERAL THEORY OF POLYADIC GROUPS	
1. Definition of a polyadic group . . . . .	213
2. Identity, inverse, equivalence . . . . .	214
3. The coset theorem . . . . .	218
4. Subgroups and transforms; expansion in cosets . . . . .	221
5. Reducibility . . . . .	228
6. Arbitrary containing ordinary groups . . . . .	238
7. Determination of all types of semi-abelianisms . . . . .	242
8. On the construction of polyadic groups . . . . .	245
II. FINITE POLYADIC GROUPS	
A. $m$ -ADIC SUBSTITUTIONS AND SUBSTITUTION GROUPS	
9. The symmetric $m$ -adic substitution group of degree $n$ . . . . .	248
10. $2^{m-1}$ -fold classification of $m$ -adic substitutions; the $m$ -adic alternating groups . . . . .	250
11. Associated and containing ordinary groups; commutative $m$ -adic substitutions . . . . .	253
12. Further study of the complete $m$ -adic $\delta$ -group and $m$ -adic alternating groups . . . . .	255
13. Transitive $m$ -adic substitution groups . . . . .	261
14. Intransitive $m$ -adic substitution groups . . . . .	262
15. Substitutions which are commutative with each of the substitutions of a transitive $m$ -adic substitution group . . . . .	263
16. Holomorphs of a regular $m$ -adic substitution group . . . . .	267
17. $m$ -adic groups of $\mu$ -adic substitutions . . . . .	272
18. Primitive and imprimitive $(m, \mu)$ substitution groups . . . . .	273
19. Multiple transitivity; cyclically transitive $m$ -adic substitution groups . . . . .	276
20. Class of an $m$ -adic substitution group . . . . .	278
B. FINITE ABSTRACT POLYADIC GROUPS	
21. Cyclic polyadic groups; ordinary theory . . . . .	282
22. Cyclic polyadic groups; polyadic theory . . . . .	286
23. Abstract polyadic groups of the first three orders . . . . .	293
24. Properties of transforms . . . . .	295
25. Generation of polyadic groups by two groups, one invariant under the elements of the other . . . . .	298
26. $m$ -adic groups of order $g$ prime to $m-1$ . . . . .	304
27. Sylow subgroups of order $p^a$ with $g/p^a$ prime to $m-1$ . . . . .	307
28. Representation of an arbitrary $m$ -adic group as a regular $m$ -adic substitution group . . . . .	312
29. Invariant subgroups and quotient groups; the $m$ -adic central quotient group . . . . .	313

Presented to the Society, October 26, 1935; received by the editors January 4, 1940.

30. Commutator, semi-commutator, and quasi-commutator subgroups . . . . .	316
31. The $\phi$ -subgroup of an $m$ -adic group . . . . .	322
32. Simply isomorphic $m$ -adic groups; group of inner isomorphisms . . . . .	324
33. Extension of Frobenius's theorem to $m$ -adic groups . . . . .	327
34. Representation of an abstract $m$ -adic group as a transitive $(m, \mu)$ substitution group . . . . .	328

C. FINITE  $m$ -ADIC LINEAR GROUPS

35. $m$ -adic linear transformations . . . . .	330
36. $m$ -adic collineations and collineation-groups . . . . .	334
37. $m$ -adic Hermitian invariants . . . . .	337
38. Reduction to canonical form . . . . .	340
39. $m$ -adic invariants . . . . .	344
40. Generalization of $m$ -adic substitution and transformation groups . . . . .	347

## INTRODUCTION

The group concept is peculiar in the breadth of its application and the narrowness of its formulation. By modifying one or more of its restrictions there have resulted such concepts as that of semi-group, groupoid, mischgruppe, quasi-group, hypergroup, multigroup. In all of these generalizations of the group concept the group operation remains dyadic, that is, it is a function of two independent variables. Our present interest is in that generalization of the group concept which results when, while retaining all other of its special features, the group operation becomes polyadic, that is, a function of any finite number of independent variables.

As far back as 1904, E. Kasner thus considered generalizing the ordinary "group property," and called a set of elements closed under a  $k$ -adic operation a  $k$ -adic system<sup>(1)</sup>. But the complete formulation of this generalization seems to have been first effected by Dörnte<sup>(2)</sup> in 1928 in a paper containing an extensive theory of what he there terms  $n$ -groups,  $n$  being the number of independent variables in the operation. In 1932 Lehmer<sup>(3)</sup> independently formulated and investigated the special concept he termed triplex, which, in Dörnte's terminology, is an abelian 3-group. Dörnte's  $m$ -group, to change

(1) While the paper in question, *An extension of the group concept*, has not appeared in print, an abstract thereof will be found in the Bulletin of the American Mathematical Society, vol. 10 (1904), pp. 290-291. Though at one point of the abstract Kasner observes that "the law of combination of the general system is best exhibited by means of its  $k$  dimensional multiplication table," his original definition adds the requirement that the combination of no fewer than  $k$  elements shall be contained in the system—a requirement that is meaningless unless the  $k$ -adic operation itself is merely an extended product based on a prior dyadic operation. And the absence of any mention of an associative law, coupled with a reference to the inverse of an element, further suggests that, as in Miller's perfect cosets referred to below, this dyadic operation is understood to be that of some actual group in the ordinary sense containing the given system.

(2) W. Dörnte, *Untersuchungen über einen verallgemeinerten Gruppenbegriff*, Mathematische Zeitschrift, vol. 29 (1928), pp. 1-19.

(3) D. H. Lehmer, *A ternary analogue of abelian groups*, American Journal of Mathematics, vol. 54 (1932), pp. 329-338.

the symbol, is also our  $m$ -adic group, or, for unspecified  $m$ , our polyadic group<sup>(4)</sup>.

As examples of triadic systems, and these also are examples of triadic groups, Kasner mentions "the odd permutations in any number of letters, the  $\infty^2$  central symmetries of the plane or the  $\infty^3$  of space, the totality of dual or reciprocal transformations, the correlations contained in any projective group, the totality of conformal transformations of the plane which reverse angles." In the introduction of his paper Dörnte mentions, among other examples, residue classes modulo  $k$  as  $(k+1)$ -groups, and in the body of his paper introduces many such arithmetical illustrations as exemplifiers of his abstract development. Apart from examples which are the subject of a major part of our theory, we may add the linear transformations of determinant an  $(m-1)$ -st root of unity as an  $m$ -group, and, more significantly, the  $m$ -group consisting of all the substitutions of a group which, instead of carrying a fixed letter into itself, transform say  $a_1 \rightarrow a_2, a_2 \rightarrow a_3, \dots, a_{m-1} \rightarrow a_1$ . In all of these examples the polyadic operation is merely an extended product expressed in terms of a prior dyadic operation. On the other hand, lengths under the operation fourth proportional, now to be written  $b:a=c:x$ , constitute a 3-group in which, geometrically, the triadic operation is primary<sup>(5)</sup>. Even more so for an abstract  $m$ -group whose operation is given ab initio by an  $m$ -dimensional table.

While the abstract formulation of polyadic group must be credited to Dörnte, in its coset theorem the present paper may be said to solve the problem of determining the essential nature of a polyadic group. This basic result is to the effect that any  $m$ -adic group can have its class of elements so widened, and in that widened class a dyadic operation so introduced, that the enlarged class, under that operation as product, constitutes an ordinary group in which the class of elements of the  $m$ -adic group is a coset of an invariant subgroup of the ordinary group, and the operation of the  $m$ -adic group the product of  $m$  elements as elements of the ordinary group<sup>(6)</sup>. At first glance this theorem seems to be identical, for finite groups, with a result of Miller's

<sup>(4)</sup> The present paper arose as a reaction to the importance ascribed to the group concept by C. J. Keyser in his *Mathematical Philosophy*, New York, 1922, Lecture XII. But see the next to the last paragraph of this introduction. We may note that an early attempt on our part to thus generalize the group concept on the basis of its fourfold characterization had failed. But on now turning to the twofold basis as given by Miller (*Finite Groups*, below, p. 52) we found generalization to be immediate.

<sup>(5)</sup> Analytically, the operation becomes  $x = (ac)/b$ , and so a variant of  $a-b+c$ , easily seen to lead to a 3-group. This last is already present in Dörnte's paper, and generalized in his Theorem 7, §1. Note that geometrically even, the binary operation multiplication can nevertheless be defined, even if secondary. The about-to-be-mentioned coset theorem shows the same situation to obtain in general.

<sup>(6)</sup> Cf. A. Suschkewitsch, *Über die Erweiterung der Semigruppe bis zur ganzen Gruppe*, Communications de la Société Mathématique de Kharkoff, (4), vol. 12 (1935), pp. 81-87.

of 1935<sup>(7)</sup>. But, apart from other differences in hypothesis, Miller obtains the coset conclusion by essentially *assuming* the given set of elements to be in an ordinary group<sup>(8)</sup>. However, as a result of the two theorems, finite polyadic group does become identical with Miller's "perfect co-set," some of whose properties he develops, provided the latter is understood to mean set of elements and polyadic operation thereon<sup>(9)</sup>.

In addition to differences in abstract development, the present paper goes beyond Dörnte's in generalizing the concepts of substitution and linear transformation in such a way that the resulting  $m$ -adic substitutions and  $m$ -adic linear transformations naturally lead to  $m$ -adic groups thereof (see §9 and §35 for their definition). These  $m$ -adic groups we study as generalizations of ordinary substitution and linear transformation groups. As incentive for this development, we have the theorem that any abstract  $m$ -adic group (finite) can be represented as a "regular"  $m$ -adic substitution group, a theorem which, indeed, first gave us our coset theorem. In the final section of the paper these concepts receive a wide extension which remains significant for ordinary groups. But they are then seen to be at least closely related to a type of ordinary group formulated by Specht<sup>(10)</sup>.

Intermediate between these generalizations of substitution group is one which includes  $m$ -adic groups of ordinary substitutions. Two of our examples given above are of this type. In this connection we may mention a work of Corral<sup>(11)</sup> referred to by Miller. With substitutions on a given finite set of letters in question, Corral calls a set of substitutions a perfect brigade if closed under the operation  $ABC$ , an imperfect brigade if closed under the operation  $AB^{-1}C$ . The former is then identical with a 3-group of ordinary substitutions, the latter with a *schar* of substitutions, *schar* in Baer's<sup>(12)</sup> wider form of a concept due to Prüfer<sup>(13)</sup>. Prüfer's development had a great influence on

(7) G. A. Miller, *Sets of group elements involving only products of more than  $n$* , Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 45-47. All references to Miller other than to *Finite Groups* (below) concern this paper.

(8) The closing statement in Kasner's abstract, which suggests an anticipation of our coset theorem for triadic groups, is more probably merely related thereto in similar fashion.

(9) His condition that his set  $S$  contain no like subset is in error. Recognizing  $S$  as an  $(n+1)$ -group of order  $h$ , we see from our §21 that his partial condition " $h$  is a power of  $n$ " should be "every distinct prime factor of  $h$  is a factor of  $n$ ."

(10) W. Specht, *Eine Verallgemeinerung der Permutationsgruppen*, Mathematische Zeitschrift, vol. 37 (1933), pp. 321-341.

(11) J. I. Corral, *Brigadas de Substituciones*, Part I, Havana, 1934; Part II, Toledo, 1935.

(12) R. Baer, *Zur Einführung des Scharbegriffs*, Journal für die reine und angewandte Mathematik, vol. 160 (1929), pp. 199-207. His abstract formulation occurs in the important footnote on page 202. (Condition III therein can be proved in its entirety, and so is unnecessary.) The same footnote proves, in our terminology (see §5), that every *schar* is reducible to an ordinary group. Had the same situation obtained for polyadic groups, there would have been no need of our coset theorem.

(13) H. Prüfer, *Theorie der Abelschen Gruppen*, I. Grundeigenschaften, Mathematische Zeitschrift, vol. 20 (1924), pp. 165-187.



Dörnte who showed that by rewriting the operation  $AB^{-1}C$  formally as  $ABC$ , Prüfer's schar becomes a special kind of 3-group. This reinterpretation is however no longer possible if the Prüfer hypothesis  $AB^{-1}C = CB^{-1}A$  is deleted to give Baer's schar.

While Dörnte's development in large measure consists in extending Prüfer's schar results to  $n$ -groups, our own work correspondingly attempts to generalize ordinary group theory. Thus, at the very beginning of our developments, where Dörnte recognizes no identity for an  $m$ -group with  $m > 2$ , we find that role played by certain sequences of  $m-1$  elements of the  $m$ -group, and are thus led to a development culminating in the coset theorem of §3. The remainder of Part I, which is really a theory of abstract polyadic groups finite or infinite, consists of largely unrelated topics, but each fundamental in the theory. Our program crystallizes in Part II which, in  $A, B, C$ , systematically generalizes most of the general topics of three chapters in the Miller, Blichfeldt, Dickson, *Finite Groups*<sup>(14)</sup>, that is, Miller's Chapters II and III on substitution groups and abstract groups respectively, and Chapter IX, Blichfeldt's introductory chapter on linear groups. The reader will find here certain developments which merely paraphrase the ordinary theory, others which are far richer in their polyadic form, and still others which have no counterpart in ordinary theory. On the whole, the amount that does go over is surprisingly large. The principal failure is the but partial extension of Sylow's theorem. To the student of ordinary groups we may point out, among other connections, that the generalizations quasi-abelianism and quasi-commutator subgroup of §30 remain significant for ordinary groups, that §5 also gives a polyadic superstructure to any ordinary group, and that the coset theorem could be used to translate polyadic group results independently arrived at into ordinary group properties. While much of Dörnte's paper becomes clarified by means of our coset theorem, and several of his developments are carried considerably further in our own work, the present paper by no means can be said to supplant Dörnte's. We are furthermore directly indebted to him for his concepts of semi-invariant subgroup and semi-abelian group.

Useful as the coset theorem is in establishing certain properties of polyadic groups, its very existence greatly minimizes the significance of that generalization. Nevertheless, we cannot agree with Miller who says "the generalization secured by using perfect cosets instead of groups is, however, only apparent." In its autonomous formulation, polyadic group is fundamentally a generalization of ordinary group and, indeed, it is as generalization that

<sup>(14)</sup> New York, 1916. Henceforth referred to as *Finite Groups*. Where in Part II the writer refers to the standard proof of an ordinary group result it is the proof in this text that is meant. We may note here that when an ordinary group term is applied without explicit definition to polyadic groups, its polyadic definition is entirely similar.

it lends itself to a corresponding development<sup>(15)</sup>. However, the final verdict will undoubtedly hang on the question of application<sup>(16)</sup>. For this end our concept of  $m$ -adic invariant is no doubt far too special (see §39). Genuine application of polyadic groups will probably therefore have to wait upon the formulation of an adequate concept of polyadic invariant.

We wish here to express our obligation to B. P. Gill to whose efforts we owe the completion of a major phase of our development (see §12). Had we completed the determination of the triadic linear groups in two variables mentioned in our preliminary report, this obligation would have been still greater. We are also indebted to R. Baer who, on two separate occasions, set us on the right path in the maze of ordinary group literature.

### I. GENERAL THEORY OF POLYADIC GROUPS

1. **Definition of a polyadic group.** Given a class of elements  $C$ , and an operation  $c(s_1 s_2 \cdots s_m)$ , we shall say that the elements of  $C$  constitute an  $m$ -adic group  $G$  under  $c$  if the following two conditions are satisfied:

1. If any  $m$  of the  $m+1$  symbols in an equation of the form

$$c(s_1 s_2 \cdots s_m) = s_{m+1}$$

represent elements in  $C$ , the remaining symbol also represents an element in  $C$ , and is uniquely determined by this equation.

2. The elements of  $C$  satisfy the associative law under  $c$ , that is, they satisfy

$$\begin{aligned} c(c(s_1 s_2 \cdots s_m) s_{m+1} s_{m+2} \cdots s_{2m-1}) &= c(s_1 c(s_2 \cdots s_m s_{m+1}) s_{m+2} \cdots s_{2m-1}) \\ &= \cdots = c(s_1 s_2 \cdots c(s_m s_{m+1} s_{m+2} \cdots s_{2m-1}))^{(17)}. \end{aligned}$$

<sup>(15)</sup> It is fundamental to remember, in this connection, that we are dealing not with a mere class of elements, but with a class of elements and an operation thereon; still better, with the properties of a class of elements under a given operation. Thus the genuineness of non-Euclidean geometry is not affected because it can be represented by certain constructions in Euclidean geometry. Had Miller's point of view been adopted, such a development as that of §5, for example, would hardly have been possible.

<sup>(16)</sup> E.g., such as the Galois theory in the case of ordinary groups, not applications, such as the examples given above, which are mere illustrations of polyadic groups or of the theory thereof. Much of Corral's development concerns a brigade Galois theory. But this seems to the writer to be merely a restatement of standard Galois theory in terms of brigades rather than a genuine application.

<sup>(17)</sup> This formulation, patterned by the author after Miller, is identical with Dörnte's except that Dörnte splits up our 1 into two parts,  $P_1$  and  $P_2$ , according as  $S_{m+1}$ , or  $S_i$ ,  $i \neq m+1$ , is to be determined. It is then readily proved by the methods of our next section that in  $P_2$  only the existence of the solution  $S_i$  need be postulated, its uniqueness being then provable. It can further be shown that this existence of a solution for  $S_i$  need only be universally postulated either for a single  $i$  with  $1 < i < m$ , or for both  $i = 1$  and  $i \neq m$ , the existence of a solution for  $S_i$  for all other  $i$ 's from 1 to  $m$  then being provable. If the second form be used in place of  $P_2$ , and the first can only be used for  $m > 2$ , the resulting set of postulates would be the exact generalization of the basis for ordinary groups used by Albert in his *Modern Higher Algebra*.

We shall also use Dörnte's phrase " $m$ -group" for  $G$ . Though these conditions are vacuously satisfied when  $C$  is a null class, the ordinary group concept tacitly assumes the existence of at least one element, and so we make the same assumption here. An ordinary group is then immediately an  $m$ -adic group with  $m=2$ , that is, a dyadic group, or 2-group. Unlike Dörnte, we exclude the case  $m=1$ .

It is readily proved by induction that the number of elements entering into any combination of elements built up by the operation  $c$  is of the form  $k(m-1)+1$ , where, in fact,  $k$  is the number of  $c$ 's in the assumed symbolic expression of this "extended operation." As the basic operation  $c(s_1s_2 \cdots s_m)$  is on an ordered  $m$ -ad of elements, an extended operation built up by  $c$ 's orders the  $k(m-1)+1$  elements appearing therein in a linear array  $s_1, s_2, \cdots, s_{k(m-1)+1}$ . It is then readily proved that as a consequence of the associative law 2 the element given by such an extended operation depends only on the sequence  $s_1, s_2, \cdots, s_{k(m-1)+1}$ , and is independent of the particular way in which parentheses are introduced in conjunction with the  $k$   $c$ 's that must enter into such an expression. We are justified, then, in briefly writing any such extended operation  $c(s_1s_2 \cdots s_{k(m-1)+1})$ .

2. **Identity, inverse, equivalence.** Let  $a_1, a_2, \cdots, a_{m-1}, a_m$  be elements of an  $m$ -adic group  $G$  satisfying the equation

$$c(a_1a_2 \cdots a_{m-1}a_m) = a_m.$$

Assuming as we do that  $m \geq 2$ , we can, in fact, let  $a_m$  and  $m-2$  of the  $m-1$  elements  $a_1, a_2, \cdots, a_{m-1}$  be arbitrary elements of  $G$ , and then determine the remaining element in accordance with 1 of §1 so that this equation will be satisfied. If now  $s$  be any element of  $G$ , we can likewise find  $s_2, s_3, \cdots, s_m$  in  $G$  so that  $c(a_ms_2s_3 \cdots s_m) = s$ . By our assumed equation we will have

$$c(c(a_1a_2 \cdots a_{m-1}a_m)s_2s_3 \cdots s_m) = c(a_ms_2s_3 \cdots s_m).$$

Hence, by the associative law,

$$c(a_1a_2 \cdots a_{m-1}c(a_ms_2s_3 \cdots s_m)) = c(a_ms_2s_3 \cdots s_m),$$

and so

$$c(a_1a_2 \cdots a_{m-1}s) = s.$$

That is, if the equation  $c(a_1a_2 \cdots a_{m-1}s) = s$  holds for one  $s$  in  $G$ , it holds for every  $s$  in  $G$ . The sequence, or  $(m-1)$ -ad,  $\{a_1, a_2, \cdots, a_{m-1}\}$  may then be called a left identity of  $G$ . In the same way we can show that if  $c(sb_1b_2 \cdots b_{m-1}) = s$  holds for one  $s$  in  $G$ , it holds for every  $s$  in  $G$ , and  $\{b_1, b_2, \cdots, b_{m-1}\}$  may be called a right identity of  $G$ .

We now prove that every left identity of  $G$  is a right identity, and conversely, thus arriving at the unique concept of an  $(m-1)$ -ad as an identity of an  $m$ -adic group. Let  $\{a_1, a_2, \cdots, a_{m-1}\}$  be a left identity. Then  $c(a_1a_2 \cdots a_{m-1}a_1) = a_1$ . By the associative law,

$$c(a_0 a_1 a_2 \cdots a_{m-2} c(a_{m-1} a_1 a_2 \cdots a_{m-1})) = c(a_0 c(a_1 a_2 \cdots a_{m-2} a_{m-1} a_1) a_2 \cdots a_{m-1}).$$

Hence

$$c(a_0 a_1 a_2 \cdots a_{m-2} c(a_{m-1} a_1 a_2 \cdots a_{m-1})) = c(a_0 a_1 a_2 \cdots a_{m-1}).$$

Since the first  $m-1$  arguments of the two members of this equation are identical, the last must also be equal by 1, §1. Hence

$$c(a_{m-1} a_1 a_2 \cdots a_{m-1}) = a_{m-1},$$

and  $\{a_1, a_2, \cdots, a_{m-1}\}$  is also a right identity. Similarly for the converse.

Our equation  $c(a_1 a_2 \cdots a_{m-1} a_1) = a_1$  shows that if  $\{a_1, a_2, \cdots, a_{m-1}\}$  is an identity, so is  $\{a_2, \cdots, a_{m-1}, a_1\}$ . Hence also  $\{a_3, \cdots, a_{m-1}, a_1, a_2\}$ , and so on. Of course we have used the preceding result on left identities being the same as right identities. In general, then, if  $\{a_1, \cdots, a_i, a_{i+1}, \cdots, a_{m-1}\}$  is an identity, so is  $\{a_{i+1}, \cdots, a_{m-1}, a_1, \cdots, a_i\}$ . Otherwise stated, cyclic permutation of the elements of an identity leaves it an identity.

Our initial observation proved the existence of an identity for  $m \geq 2$ . Clearly, if  $\{a_1, a_2, \cdots, a_{m-1}\}$  is an identity, it is immaterial which  $m-2$  of these elements were assumed arbitrarily. Hence all identities of an  $m$ -adic group can be obtained by arbitrarily assigning values to, say,  $a_1, a_2, \cdots, a_{m-2}$ , and correspondingly determining  $a_{m-1}$ . If  $G$  be of finite order  $g$ , there are  $g^{m-1}$   $(m-1)$ -ads formed from elements of  $G$ . Hence  $G$  has  $g^{m-2}$  identities. There will be no ambiguity if we use similar terminology when  $g$  is infinite.

While the term identity will thus mean an  $(m-1)$ -ad of the above kind, a corresponding development in connection with an extended operation on  $k(m-1)+1$  arguments leads to what may be termed an extended identity in the form of a  $k(m-1)$ -ad. Except for their number, extended identities enjoy the same properties as identities. Rather unsymmetrically we may say that  $\{a_1, a_2, \cdots, a_{k(m-1)}\}$  is an extended identity if  $\{a_1, a_2, \cdots, a_{m-2}, c(a_{m-1} \cdots a_{k(m-1)})\}$  is an identity.

The concept of identity immediately leads to that of inverse. For  $m=2$ , the inverse of an element  $s$  is an element which multiplied into  $s$  yields the identity. For  $m>2$ , to obtain an identity from an element  $s$  we must annex  $m-2$  other elements. We are thus led to an  $(m-2)$ -ad as an inverse of  $s$ . Hence, for  $m>2$ , an inverse of an element is an element when and only when  $m=3$ .  $\{s_1, s_2, \cdots, s_{m-2}\}$  is then an inverse of  $s$  if  $\{s, s_1, s_2, \cdots, s_{m-2}\}$  is an identity. As  $\{s_1, s_2, \cdots, s_{m-2}, s\}$  is then also an identity, we may therefore say that  $s$  is an inverse of the  $(m-2)$ -ad  $\{s_1, s_2, \cdots, s_{m-2}\}$ . We are thus led to define inverse for  $i$ -ads with arbitrary  $i$ .

First let  $i < m-1$ . We then define an inverse of an  $i$ -ad  $\{s_1, s_2, \cdots, s_i\}$  to be an  $(m-i-1)$ -ad  $\{s'_1, s'_2, \cdots, s'_{m-i-1}\}$  such that  $\{s_1, s_2, \cdots, s_i, s'_1, s'_2, \cdots, s'_{m-i-1}\}$  is an identity. As  $\{s'_1, s'_2, \cdots, s'_{m-i-1}, s_1, s_2, \cdots, s_i\}$  is then also an identity,  $\{s_1, s_2, \cdots, s_i\}$  is an inverse of  $\{s'_1, s'_2, \cdots, s'_{m-i-1}\}$ , so that we can talk of a pair of inverse polyads. When  $i=m-1$  we must

have recourse to an extended identity, and are thus led to an  $(m-1)$ -ad as inverse.  $\{s'_1, s'_2, \dots, s'_{m-1}\}$  is then an inverse of  $\{s_1, s_2, \dots, s_{m-1}\}$  if  $\{s_1, s_2, \dots, s_{m-1}, s'_1, s'_2, \dots, s'_{m-1}\}$  is an extended identity. As before,  $\{s_1, s_2, \dots, s_{m-1}\}$  is also an inverse of  $\{s'_1, s'_2, \dots, s'_{m-1}\}$ .

By means of inverses we easily solve an equation of the form

$$c(a_1 a_2 \dots a_i s b_1 b_2 \dots b_{m-i-1}) = s_0$$

for  $s^{(18)}$ . Let  $\{a'_1, a'_2, \dots, a'_{m-i-1}\}$ ,  $\{b'_1, b'_2, \dots, b'_i\}$  be inverses of  $\{a_1, a_2, \dots, a_i\}$ ,  $\{b_1, b_2, \dots, b_{m-i-1}\}$  respectively. Operating on both sides of the above equation by  $c(a'_1 a'_2 \dots a'_{m-i-1} | b'_1 b'_2 \dots b'_i)$ , the bar indicating the missing argument, applying the associative law, and reducing the left-hand side by the property of identities we obtain

$$s = c(a'_1 a'_2 \dots a'_{m-i-1} s_0 b'_1 b'_2 \dots b'_i).$$

When  $a$ 's or  $b$ 's are missing, our inverse of an  $(m-1)$ -ad serves the same purpose. Clearly an equation of the same type arising from an extended operation can always be reduced to the above type by means of the associative law. Our need of inverses of  $i$ -ads with  $i > m-1$  is thus not pressing. However, they can be similarly introduced by means of extended identities. While such an inverse can always be a  $j$ -ad with  $1 \leq j \leq m-1$ , to preserve the symmetry of the inverse relationship we must allow  $j > m-1$  as well, and thus have to introduce extended inverses. Thus if  $i = k(m-1) + l$ ,  $0 \leq l < m-1$ , an inverse will be an  $(m-l-1)$ -ad, while all extended inverses will have  $j$  in the form  $\kappa(m-1) + (m-l-1)$ .

The multiplicity of inverses when the latter are not single elements leads to the concept of equivalent  $i$ -ads. We can introduce that concept directly, however, as follows. Let  $\{a_1, a_2, \dots, a_i\}$  and  $\{b_1, b_2, \dots, b_i\}$  be such that for some specific  $d_1, \dots, d_j, e_1, \dots, e_{m-i-j}$

$$c(d_1 \dots d_j a_1 a_2 \dots a_i e_1 \dots e_{m-i-j}) = c(d_1 \dots d_j b_1 b_2 \dots b_i e_1 \dots e_{m-i-j});$$

that is, replacing the sequence  $a_1, a_2, \dots, a_i$  by  $b_1, b_2, \dots, b_i$  in the specific operation given by the left-hand member of this equation leaves the result unaltered. Let  $\{d'_1, \dots, d'_{m-j-1}\}$  and  $\{e'_1, \dots, e'_{i+j-1}\}$  be inverses of  $\{d_1, \dots, d_j\}$ ,  $\{e_1, \dots, e_{m-i-j}\}$  respectively, and let  $s_1, \dots, s_\kappa, s_{\kappa+1}, \dots, s_{m-i}$

<sup>(18)</sup> Dörnte solves this equation for  $m > 2$  by his "querelement"  $\bar{a}$ , defined as the solution of the equation  $c(a \dots ax) = a$  for  $x$ . The very economy of this concept, however, helps obscure the concepts of our present section, so necessary for the basic coset theorem. It may be pointed out that actually our method of solution can be so presented as to be independent of the previous theorems on identities, and thus leads to that part of the footnote to §1 concerning the provability of the uniqueness of the solution. Indeed, in this primordial form, the same method is constantly used by Dörnte without specific formulation. The reader may be interested in noting that Dörnte's Theorems 3 and 4, §1, may be considered special cases of our identity results in that the definition of  $\bar{a}$  may now be restated:  $\{a, \dots, a, a\}$  is a right identity.



be arbitrary elements of  $G$ . Operating on both sides of the above equation by the extended operation  $c(s_1 \cdots s_k d'_1 \cdots d'_{m-j-1} | e'_1 \cdots e'_{i+j-1} s_{k+1} \cdots s_{m-i})$  we obtain, after simplification,

$$c(s_1 \cdots s_k a_1 a_2 \cdots a_i s_{k+1} \cdots s_{m-i}) = c(s_1 \cdots s_k b_1 b_2 \cdots b_i s_{k+1} \cdots s_{m-i}).$$

A similar argument can be given when  $j$ , or  $m-i-j$ , is 0 or  $m-1$ . Hence, if the sequence  $b_1 b_2 \cdots b_i$  can replace  $a_1 a_2 \cdots a_i$  somewhere in one operation it can do so anywhere in any operation<sup>(19)</sup>. Clearly the same result holds good for extended operations as well. The  $i$ -ads  $\{a_1, a_2, \cdots, a_i\}$  and  $\{b_1, b_2, \cdots, b_i\}$  will then be said to be *equivalent*. Thus we may define an  $m$ -group  $G$  to be abelian if the dyads  $\{s_1, s_2\}$  and  $\{s_2, s_1\}$  are equivalent for every pair of elements  $s_1, s_2$  of  $G$ . For then the value of  $c(s_1 s_2 \cdots s_m)$ ,  $s$ 's in  $G$ , is unaltered by any interchange of adjacent  $s$ 's, and hence by any permutation of all the  $s$ 's.

Let  $\{a_1, a_2, \cdots, a_i\}$  and  $\{b_1, b_2, \cdots, b_i\}$  be equivalent  $i$ -ads, and let  $\{a'_1, a'_2, \cdots, a'_{m-i-1}\}$  be an inverse of  $\{a_1, a_2, \cdots, a_i\}$ . We have then  $c(a'_1 a'_2 \cdots a'_{m-i-1} a_1 a_2 \cdots a_i s) = s$ . Hence also  $c(a'_1 a'_2 \cdots a'_{m-i-1} b_1 b_2 \cdots b_i s) = s$  so that  $\{a'_1, a'_2, \cdots, a'_{m-i-1}\}$  is an inverse of  $\{b_1, b_2, \cdots, b_i\}$  as well. A similar argument applies when  $i = m-1$ . That is, every inverse of one of a pair of equivalent  $i$ -ads is also an inverse of the other. Again, let  $\{a_1, a_2, \cdots, a_i\}$  and  $\{b_1, b_2, \cdots, b_i\}$  both be inverses of  $\{a'_1, a'_2, \cdots, a'_{m-i-1}\}$ . Since we then have  $c(a'_1 a'_2 \cdots a'_{m-i-1} a_1 a_2 \cdots a_i s) = s = c(a'_1 a'_2 \cdots a'_{m-i-1} b_1 b_2 \cdots b_i s)$ , it follows that  $\{a_1, a_2, \cdots, a_i\}$  and  $\{b_1, b_2, \cdots, b_i\}$  are equivalent. That is, *inverses of the same polyad are equivalent*. It follows from these results that if  $\{a'_1, a'_2, \cdots, a'_{m-i-1}\}$  is an inverse of  $\{a_1, a_2, \cdots, a_i\}$ , the class of inverses of  $\{a_1, a_2, \cdots, a_i\}$  is the class of  $(m-i-1)$ -ads equivalent to  $\{a'_1, a'_2, \cdots, a'_{m-i-1}\}$ . Conversely, the class of  $i$ -ads equivalent to  $\{a_1, a_2, \cdots, a_i\}$  is the class of inverses of  $\{a'_1, a'_2, \cdots, a'_{m-i-1}\}$ . Finally, the first class is the class of inverses of each member of the second, and conversely. This for  $i < m-1$ . For  $i = m-1$  both classes consist of  $(m-1)$ -ads.

We shall speak of the class of all  $i$ -ads equivalent to a given  $i$ -ad as a *class of equivalent  $i$ -ads*. As in the case of identities, to obtain all  $i$ -ads equivalent to a given  $i$ -ad we may assign arbitrary values to  $i-1$  of the elements, the  $i$ th being then determined. We may therefore say that a class of equivalent  $i$ -ads has  $g^{i-1}$  members. If, on the other hand, we keep  $i-1$  elements fixed, and let the remaining element run through  $G$ , 1 of §1 shows that no two of the resulting  $i$ -ads can be equivalent, while each class of equivalent  $i$ -ads thus finds a representative. We may therefore say that for each  $i$  there are exactly  $g$  classes of equivalent  $i$ -ads. These classes are, of course, mutually exclusive. For  $i=1$  they are nothing more than the unit classes consisting of single elements of  $G$ . For  $i=m-1$  one class of equivalent  $i$ -ads is singled out, that is, the class of identities.

<sup>(19)</sup> This result is proved in part by Dörnte as Theorem 2, §1, but the corresponding concept is not formulated. Clearly this relationship between  $i$ -ads is an "equivalence relationship."



3. **The coset theorem.** We are now in a position to embed our  $m$ -adic group  $G$  in an ordinary group. Let  $C^*$  be the class of all classes of equivalent  $i$ -ads for  $i=1, 2, \dots, m-1$ . Each element of  $C^*$  is thus a class of equivalent  $i$ -ads, and  $C^*$  may then be said to have  $(m-1)g$  elements,  $g$  for each  $i$ . It is convenient to drop the distinction between a unit class and its sole member, so that we may consider  $C$ , the class of elements of  $G$ , a subclass of  $C^*$ . We proceed now to define a dyadic operation on the elements of  $C^*$ . But first we must remove the above tacit restriction  $i < m$  in our discussion of equivalence. Clearly, by using extended operations, our results go over for  $i \geq m$ . Furthermore, we can extend the concept of equivalence to allow an  $i$ -ad to be equivalent to a  $j$ -ad. With only the basic operation  $c$  involved, we must clearly have  $j-i$  a multiple of  $m-1$ . Without further elaboration,  $\{b_1, b_2, \dots, b_{i+k(m-1)}\}$  will be equivalent to  $\{a_1, a_2, \dots, a_i\}$  if  $\{b_1, b_2, \dots, b_{i-1}, c(b_i \dots b_{i+k(m-1)})\}$  and  $\{a_1, a_2, \dots, a_i\}$  are equivalent in the original sense<sup>(20)</sup>.

We first prove the following: if two of the three polyads  $\{a_1, a_2, \dots, a_i\}$ ,  $\{b_1, b_2, \dots, b_j\}$ ,  $\{a_1, a_2, \dots, a_i, b_1, b_2, \dots, b_j\}$  are respectively equivalent to the corresponding two of the three polyads  $\{a'_1, a'_2, \dots, a'_i\}$ ,  $\{b'_1, b'_2, \dots, b'_j\}$ ,  $\{a'_1, a'_2, \dots, a'_i, b'_1, b'_2, \dots, b'_j\}$ , the remaining polyads are equivalent. We shall prove this result for  $i+j \leq m$ , a corresponding proof with the use of extended operations serving for  $i+j > m$ . Consider then the operations  $c(a_1 a_2 \dots a_i b_1 b_2 \dots b_j d_1 \dots d_{m-i-j})$  and  $c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j})$ . If the first and second polyads of the first set of three are respectively equivalent to the first and second of the second set of three, we will have  $c(a_1 a_2 \dots a_i b_1 b_2 \dots b_j d_1 \dots d_{m-i-j}) = c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j}) = c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j})$ , and the third polyads are equivalent. If the hypothesis concerns the first and third polyads, then  $c(a_1 a_2 \dots a_i b_1 b_2 \dots b_j d_1 \dots d_{m-i-j}) = c(a'_1 a'_2 \dots a'_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j}) = c(a_1 a_2 \dots a_i b'_1 b'_2 \dots b'_j d_1 \dots d_{m-i-j})$ , whence the corresponding conclusion. Similarly for the second and third polyads.

Let then the dyadic operation  $c^*(r_1 r_2)$  be defined as follows. If  $r_1$  and  $r_2$  are members of  $C^*$ , and if  $\{a_1, a_2, \dots, a_i\}$  is in the class  $r_1$  of equivalent  $i$ -ads,  $\{b_1, b_2, \dots, b_j\}$  in the class  $r_2$  of equivalent  $j$ -ads, then  $c^*(r_1 r_2)$  is to be the class of  $(i+j)$ -ads equivalent to  $\{a_1, a_2, \dots, a_i, b_1, b_2, \dots, b_j\}$  when  $i+j \leq m-1$ , the class of  $(i+j-(m-1))$ -ads equivalent to  $\{a_1, a_2, \dots, a_i,$

<sup>(20)</sup> And, of course, our basic theorem on equivalent  $i$ -ads extends to equivalent polyads. It may then be noted that if we include a null sequence in this framework, an independent proof of the identity of left and right identities results. In fact, the about-to-be-proved coset theorem depends only on the concept of equivalence; and the properties of identity and inverse could therefore be derived with the help of that theorem. Their direct formulation in terms of the operation of the  $m$ -group, however, will be found indispensable for correct thinking on such topics as those of §5.

$b_1, b_2, \dots, b_j\}$  when  $i+j > m-1$ . When  $i+j \leq m-1$ , our previous results not only show that  $c^*(r_1 r_2)$  is independent of the particular  $i$ -ad and  $j$ -ad chosen from  $r_1$  and  $r_2$  respectively, but that if any two symbols in the equation  $c^*(r_1 r_2) = r_3$  are assigned values in  $C^*$ , the third is uniquely determined in  $C^*$ . The same is true when  $i+j > m-1$  by the transitive property of equivalence. Hence, condition 1 of §1 for a dyadic group is satisfied by  $(C^*, c^*)$ ; likewise condition 2, that is, the associative law. For let  $\{a_1, \dots, a_i\}$ ,  $\{b_1, \dots, b_j\}$ ,  $\{c_1, \dots, c_k\}$  be in  $r_1, r_2, r_3$  respectively. Then, with equivalence extended as above, if  $i+j+k = l+\lambda(m-1)$ ,  $1 \leq l \leq m-1$ , both  $c^*(c^*(r_1 r_2) r_3)$  and  $c^*(r_1 c^*(r_2 r_3))$  represent the class of  $l$ -ads equivalent to  $\{a_1, \dots, a_i, b_1, \dots, b_j, c_1, \dots, c_k\}$ , so that, for all members of  $C^*$ ,

$$c^*(c^*(r_1 r_2) r_3) = c^*(r_1 c^*(r_2 r_3)).$$

Hence, the members of  $C^*$  constitute an ordinary group under  $c^*$ . With  $G$  as the given  $m$ -adic group, this ordinary group will be symbolized  $G^*$ .

We have observed that we may consider the members of  $G$  to be members of  $G^*$ , that is, those classes of equivalent  $i$ -ads for which  $i=1$ . We now further observe that the operation  $c(s_1 s_2 \dots s_m)$  can be identified with the extended operation  $c^*(s_1 s_2 \dots s_m)$  when, of course, the  $s$ 's are in  $G$ . For  $c^*(s_1 s_2 \dots s_m)$  is, indeed, the class of monads equivalent to  $\{s_1, s_2, \dots, s_m\}$ , and so consists of but the one monad  $c(s_1 s_2 \dots s_m)^{(21)}$ . We shall therefore call  $G^*$  the *abstract containing ordinary group of  $G$* , abstract by contrast with other possibilities to be discussed later. In fact,  $G^*$  is clearly determined by the abstract form of  $G$ . And while  $G^*$  as derived is not abstract, it may be made so by replacing the members of  $C^*$  by symbols formally obeying the rule of combination  $c^*$  as determined above.

To obtain a clearer view of the relationship between  $G$  and  $G^*$ , and thus, indeed, really to solve the problem of the essential nature of a polyadic group, let us consider those members of  $G^*$  which are classes of equivalent  $(m-1)$ -ads. We have already observed that one of these  $g$  classes is the class of identities of  $G$ . Now if in the equation

$$c^*(r_1 r_2) = r_3$$

any two of the three symbols represent classes of equivalent  $(m-1)$ -ads, so does the third. It follows that the  $g$  classes of equivalent  $(m-1)$ -ads constitute an ordinary group under  $c^*$ , and hence a subgroup of  $G^*$ . We shall symbolize this ordinary group by  $G_0$ , and call it the *associated ordinary group of  $G$* . It is readily seen that  $G_0$  is an invariant subgroup of  $G^*$ <sup>(22)</sup>. To prove that, it

<sup>(21)</sup> If then  $G$  has but a finite number of elements, Miller's theorem concerning perfect cosets can be applied immediately to give the coset theorem that follows. However we here make no such restriction on  $G$ .

<sup>(22)</sup> Provided  $m > 2$ . For  $m=2$ ,  $G^*=G=G_0$ . If then we here allow the term subgroup to include the group itself, the results of the present section are also valid for ordinary groups, though in trivial fashion.

is sufficient to show that in the equation

$$c^*(tr_1) = c^*(r_1r_2)$$

if  $t$  is in  $G_0$ ,  $r_1$  in  $G^*$ , then  $r_2$  is in  $G_0$ . But if  $r_1$  is a class of equivalent  $i$ -ads,  $t$  being a class of equivalent  $(m-1)$ -ads, then  $c^*(tr_1)$ , and hence  $c^*(r_1r_2)$ , is also a class of equivalent  $i$ -ads.  $r_2$ , then, can only be a class of equivalent  $(m-1)$ -ads, as was to be proved.

Let us now expand  $G^*$  in cosets as regards its invariant subgroup  $G_0$ . As in the invariance proof, if a multiplier  $r$  represents a class of equivalent  $i$ -ads, the corresponding coset consists of classes of equivalent  $i$ -ads, and indeed, constitutes the class of all  $g$  classes of equivalent  $i$ -ads. While this is immediate when  $g$  is finite, in any case if  $r_1$  is a class of equivalent  $i$ -ads, the equation  $c^*(r_2r) = r_1$  demands that  $r_2$  be in  $G_0$ , so that  $r_1$  is in the coset in question. Hence the expansion of  $G^*$  as regards  $G_0$  consists of exactly  $m-1$  augmented cosets, each being the class of all  $g$  classes of equivalent  $i$ -ads, for some  $i = 1, 2, \dots, m-1$ . The elements of  $G$  itself therefore constitute one of these cosets, that is, that one for which  $i = 1$ . Hence our basic theorem. *Every polyadic group is a coset of an ordinary group with respect to an invariant subgroup*, it being understood that the polyadic operation of the polyadic group is an extension of the dyadic operation of the ordinary group.

With the relationship between  $G$ ,  $G_0$  and  $G^*$  made thus precise, it becomes desirable to simplify our notation. Hence, when but a single  $m$ -adic operation  $c$  is involved, we shall write the corresponding dyadic operation  $c^*(r_1r_2)$  simply as the product  $r_1r_2$  of standard group theory. Our identification of  $c(s_1s_2 \dots s_m)$  with  $c^*(s_1s_2 \dots s_m)$  therefore enables us to write  $c(s_1s_2 \dots s_m)$ , simply,  $s_1s_2 \dots s_m$ . We now finally introduce the completely abstract view of  $G^*$  with symbols for elements. Clearly the element of  $G^*$  corresponding to the class of identities of  $G$  is the identity of  $G^*$ , and so will be symbolized by 1, as usual. With the elements of  $G^*$  as symbols, it will be convenient to call the symbol  $r$ , representing a class of equivalent  $i$ -ads, an  $i$ -ad. Thus  $s_1s_2 \dots s_i$  will be an  $i$ -ad when the  $s$ 's are elements of  $G$ . Conversely, every  $i$ -ad can be written thus. In particular,  $G_0$ , itself, consists of all distinct products  $s_1s_2 \dots s_{m-1}$  of  $m-1$  elements in  $G$ . To avoid duplication, of course, we may keep  $m-2$  of these elements fixed, and let the remaining one run through  $G$ .

In particular, if  $s$  is an element of  $G$ ,  $s^i$  is an  $i$ -ad, and so may correspondingly be used as multiplier in the expansion of  $G^*$  in cosets as regards  $G_0$ . We may therefore write this expansion

$$G^* = G_0s + G_0s^2 + \dots + G_0s^{m-2} + G_0 = sG_0 + s^2G_0 + \dots + s^{m-2}G_0 + G_0.$$

Most significantly we may then also write

$$G = G_0s = sG_0.$$

Since  $G_0$  consists of products of elements of  $G$ , we see that  $G^*$  itself is generated by the elements of  $G$ . The expansion of  $G^*$  shows the quotient group  $G^*/G_0$  to be of order  $m-1$ , and, indeed, cyclic, with the element corresponding to the given polyadic group  $G$  as generator. Our *coset theorem* is thus more precise than its brief formulation, given above, would indicate.

By means of this theorem we shall be able to prove many results concerning polyadic groups by means of known results on ordinary groups. On the other hand, the following almost immediately obvious converse enables polyadic group theory to make contributions to a certain aspect of ordinary group theory. To wit, *if a coset of an ordinary group with respect to an invariant subgroup is of finite order  $m-1$  as element of the corresponding quotient group, then the elements of the coset constitute a polyadic group under the product of  $m$  elements as operation*<sup>(2)</sup>. Though easily proved directly, this result may be considered a consequence of the general theorem of §8. It will also be generalized at the end of the next section. Note that such a result cannot be true for a coset corresponding to an element of infinite order of the quotient group.

**4. Subgroups and transforms; expansion in cosets.** Dörnte has treated the subject of expansions of polyadic groups in cosets exhaustively. While not possessing identities and inverses to lead to a concept of transforms, he was enabled adequately to treat invariant subgroups by mere commutativity properties. He further introduced what we shall refer to as semi-invariant subgroups, a concept which the writer completely overlooked in his own development, and was thus led to a more general concept of polyadic quotient groups than is given by invariant subgroups. Nevertheless we shall reexamine these concepts from the point of view of the coset theorem, and a theory of transforms, since not only do they become clearer thereby, but indeed admit of a certain degree of generalization.

A proper subclass of the class of elements of an  $m$ -adic group  $G$  will be said to constitute a subgroup  $H$  of  $G$  if the elements of that subclass constitute a polyadic group under the polyadic operation of  $G$ . This is clearly equivalent to the following. If in an equation  $c(s_1 s_2 \cdots s_m) = s_{m+1}$  any  $m$  elements are in the subclass, the  $(m+1)$ -st is. For the rest of the definition of  $m$ -adic group follows from the elements of the subclass being in  $G$ . Where no confusion can result we shall occasionally allow  $G$  to be a subgroup (improper) of itself. We proceed first to investigate the relationship between  $H^*$  and  $G^*$ ,  $H_0$  and  $G_0$ .

With  $H^*$  and  $G^*$  considered as being composed of classes of equivalent  $i$ -ads, only those members of  $H^*$  which are in  $H$  will also be members of  $G^*$ . For if  $\{s_1, s_2, \cdots, s_i\}$  is an  $i$ -ad of  $H$ , and hence also of  $G$ , the class of  $H$   $i$ -ads equivalent to  $\{s_1, s_2, \cdots, s_i\}$  is but a proper subclass of the class of  $G$   $i$ -ads equivalent to  $\{s_1, s_2, \cdots, s_i\}$  whenever  $i > 1$ . Nevertheless a 1-1 correspondence is thus set up between the members of  $H^*$  and the members of  $G^*$

(2) Already proved by Miller in equivalent form for finite groups.

containing them. For the latter are mutually exclusive. Hence, when  $G^*$  is treated abstractly with symbols as elements, we may symbolize the members of  $H^*$  correspondingly; and as the operation  $c^*(s_1s_2)$ , that is,  $s_1s_2$  as explained above, when set up for  $G^*$  now serves also for  $H^*$ ,  $H^*$  thereby becomes a subgroup of  $G^*$ .

The  $(m-1)$ -ads of  $H^*$  are then also  $(m-1)$ -ads of  $G^*$ , so that  $H_0$  is a subgroup of  $G_0$ . If  $s$  is any element of  $H$ , we can simultaneously expand  $H^*$  and  $G^*$  in the form

$$H^* = H_0s + H_0s^2 + \cdots + H_0s^{m-2} + H_0,$$

$$G^* = G_0s + G_0s^2 + \cdots + G_0s^{m-2} + G_0.$$

It follows that the  $m-1$  augmented cosets of  $H^*$  as regards  $H_0$  are respectively contained in the  $m-1$  augmented cosets of  $G^*$  as regards  $G_0$ . As an immediate consequence, we have *Lagrange's theorem holds for finite polyadic groups*. For, defining the order of a polyadic group as the number of its elements, the relations  $G = G_0s$ ,  $H = H_0s$  show that the order  $g$  of the polyadic group  $G$ , and the order  $h$  of its subgroup  $H$ , are respectively the same as the order of the ordinary group  $G_0$ , and its subgroup  $H_0$ ; and hence,  $h$  is a divisor of  $g$ .

Since  $H$  generates  $H^*$ , and in turn consists of the common elements of  $H^*$  and  $G$ , the correspondence between the subgroups  $H$  of  $G$ , and their abstract containing groups  $H^*$ , is 1-1.  $H_0$  consists of the common elements of  $H^*$  and  $G_0$ , and hence is also determined by  $H$ . In fact, we shall find useful the result that the products of  $m-1$  elements chosen from a subgroup  $H$  of  $G$  constitute a subgroup of  $G_0$ , namely  $H_0$ . On the other hand, different subgroups of  $G$  may have the same associated ordinary group  $H_0$ . Hence, in general, we can only say that the correspondence between the subgroups  $H$  of  $G$ , and their associated ordinary groups  $H_0$ , is but many-one. Furthermore, not every subgroup  $H_0$  of  $G_0$  need be the associated ordinary group of a subgroup  $H$  of  $G$ . The coset theorem and its converse, indeed, show that *the necessary and sufficient condition that a subgroup  $H_0$  of  $G_0$  be the associated ordinary group of some subgroup  $H$  of  $G$  is that there exist an element  $s$  of  $G$  such that  $H_0$  is invariant under  $s$ , while  $s^{m-1}$  is in  $H_0$* . Indeed the subgroups of  $G$  are the distinct  $H_0s$ 's obtained from all  $H_0$ 's and  $s$ 's satisfying this condition.

As has been observed by Dörnte, two subgroups  $H$  and  $K$  of an  $m$ -adic group  $G$  need have no element in common. Thus, this will always be so if  $H$  and  $K$  are distinct subgroups of  $G$  with the same associated group. If, however,  $H$  and  $K$  do have an element in common, their common elements clearly constitute a subgroup of each of the subgroups, if they are not identical with one or the other. Moreover, if  $s$  be such a common element, by writing  $H = H_0s$ ,  $K = K_0s$ , we see that the associated group of the "crosscut" of  $H$  and  $K$  is the crosscut of their associated groups.



We consider next the expansion of  $G$  in cosets as regards a subgroup  $H$  thereof.  $H_0$  is clearly a subgroup of  $G^*$ . We may therefore expand  $G^*$  in say right cosets as regards  $H_0$ . Now it is immediately seen that such a coset of  $H_0$  either has no element in  $G$ , or is completely contained in  $G$ . For if this coset has an element  $s$  in common with  $G$ , then, since the coset can be written  $H_0s$ , and  $H_0$  is contained in  $G_0$ ,  $H_0s$  will be wholly contained in  $G = G_0s$ . As all the elements of  $G$  must appear in the given expansion of  $G^*$ , we see that the cosets in question containing elements of  $G$  constitute a separation of the elements of  $G$  into mutually exclusive classes of elements. We may say then that  $G$  has thus been *expanded in right cosets as regards  $H$* . A similar result holds for *left cosets*.

And now an immediate generalization. In the above discussion  $H$  served only to introduce the subgroup  $H_0$  of  $G_0$ . If then  $H_0$  be any subgroup of  $G_0$ , whether it corresponds to a subgroup  $H$  of  $G$ , or not, the above argument holds without change. Hence, *every subgroup of the associated ordinary group of a polyadic group leads to an expansion of the polyadic group in right cosets, and in left cosets, as regards that subgroup.*

Specifically, if in the expansion of  $G^*$  in right cosets as regards  $H_0$  the corresponding multipliers which are in  $G$  are  $s_\alpha, s_\beta, \dots, s_\epsilon$ , then the expansion of  $G$  in right cosets as regards  $H_0$  can be written

$$G = H_0s_\alpha + H_0s_\beta + \dots + H_0s_\epsilon.$$

Similarly for left cosets. A not easily proved theorem for ordinary finite groups is that the coset multipliers may be so selected that they are the same on the right as on the left. An immediate corollary of the preceding formulation is that the same is true of finite polyadic groups.

It is sometimes necessary to consider the intersections of cosets in the expansion of  $G$  in, say, right cosets as regards subgroups  $H_0$ , and  $K_0$ , of  $G_0$ . We have then immediately that while a coset with respect to  $H_0$  and a coset with respect to  $K_0$  may have no elements in common, if they do have a common element  $s$ , then their common elements constitute the set  $L_0s$  where  $L_0$  is the crosscut of  $H_0$  and  $K_0$ . In particular, if  $G$  is finite, all such intersecting pairs of cosets intersect in the same number of elements, namely, a number equal to the order of the crosscut of  $H_0$  and  $K_0$ .

Expansions of  $G$  in double cosets likewise admit of simple treatment. With  $H_0$  and  $K_0$  arbitrary subgroups of  $G_0$ , we may expand  $G^*$  in double cosets  $H_0rK_0$ . If any element of such a double coset is in  $G$ , the entire double coset is contained in  $G$ . Hence, if in the expansion of  $G^*$  we select those double cosets with  $r$  in  $G$ , the result will be a separation of the elements of  $G^*$  into mutually exclusive sets, that is, the expansion of  $G$  in double cosets as regards  $H_0$  and  $K_0$ . In particular, if  $G$  has subgroups  $H$  and  $K$  whose associated ordinary groups are  $H_0$  and  $K_0$  respectively, the resulting expansion may be



spoken of as the expansion of  $G$  in double cosets as regards  $H$  and  $K$ , the case considered by Dörnte<sup>(24)</sup>.

We shall introduce the property of invariance through the more general concept of transform. To insure the fundamental correctness of our concept, we go back to first principles. Given an element  $s$ , and an  $i$ -ad  $\{s_1, s_2, \dots, s_i\}$ , both considered in the  $m$ -adic sense, we define the *transform* of  $s$  under  $\{s_1, s_2, \dots, s_i\}$  to be the element

$$c(s'_1 s'_2 \dots s'_{m-i-1} s s_1 s_2 \dots s_i)$$

where  $\{s'_1, s'_2, \dots, s'_{m-i-1}\}$  is an inverse of  $\{s_1, s_2, \dots, s_i\}$ . This for  $i < m-1$ ; a similar definition holds for  $i = m-1$ . Since all inverses of a given polyad are equivalent, this transform is uniquely determined by  $s$ , and  $\{s_1, s_2, \dots, s_i\}$ . Since inverses of equivalent  $i$ -ads are also equivalent, it follows that equivalent  $i$ -ads yield identical transforms of a given element.

In saying  $s$  and  $\{s_1, s_2, \dots, s_i\}$  are  $m$ -adic, we tacitly assume that there is some  $m$ -adic group to which  $s, s_1, s_2, \dots, s_i$  belong. Let us then consider the abstract containing ordinary group of this  $m$ -adic group, and treat it in abstract form, with simplified notation. If, then,  $i$ -ad  $\{s_1, s_2, \dots, s_i\}$  corresponds to abstract  $i$ -ad  $r$  of the containing group, the  $(m-i-1)$ -ad  $\{s'_1, s'_2, \dots, s'_{m-i-1}\}$  will correspond to an abstract  $(m-i-1)$ -ad  $r'$  such that if  $s$  be an element of the  $m$ -adic group,  $r's = s$ . Writing the identity of the containing group as usual, we thus have  $r'r = 1$ , and hence in customary notation,  $r' = r^{-1}$ . Consequently, if  $r$  represents a class of equivalent polyads of a polyadic group,  $r^{-1}$  represents the class of inverses of those polyads. The transform of  $s$  under  $\{s_1, s_2, \dots, s_i\}$  can now be written  $r^{-1}sr$ . And so, the transform of an element by an  $i$ -ad is the ordinary transform of that element by the corresponding abstract  $i$ -ad in the abstract containing group.

We can now extend our concept of transform to that of the transform of a polyad by a polyad. In general, via the abstract containing group, the transform of  $r_1$  by  $r_2$  is  $r_2^{-1}r_1r_2$ . Had we resorted to our primitive concepts in this case, we would have, as with inverses, a class of equivalent transforms. We readily see that in all cases the transform of an  $i$ -ad,  $i \leq m-1$ , is an  $i$ -ad.

Consider now an  $m$ -adic group  $G$ , and an  $i$ -ad  $r$  not necessarily an  $i$ -ad of  $G$ . Then, as with ordinary groups, if each element of  $G$  is transformed by  $r$ , there results an  $m$ -adic group  $G'$  which may be said to be simply isomorphic with  $G$ , and will be termed the transform of  $G$  under  $r$ . In fact, let  $s'$  be the transform under  $r$  of any element  $s$  of  $G$ . Since  $r^{-1}s_1r \cdot r^{-1}s_2r \cdot \dots \cdot r^{-1}s_mr = r^{-1}s_1s_2 \cdot \dots \cdot s_mr$ , we see that the relationship  $s_1s_2 \cdot \dots \cdot s_m = s_{m+1}$  is equivalent

<sup>(24)</sup> At first glance it would appear that Dörnte's expansions in cosets and double cosets, while depending on actual subgroups of  $G$ , are more general than we have stated them to be. However, it is readily seen that Dörnte's expansions with respect to a subgroup, or subgroups, of  $G$  are our expansions of  $G$  with respect to transforms, in the sense defined below, of the given subgroup or subgroups by polyads of  $G$ . And since these transforms are again subgroups of  $G$ , the Dörnte expansions are no more general than we have stated them to be.

to  $s'_1 s'_2 \cdots s'_m = s'_{m+1}$ . The defining properties 1 and 2 for an  $m$ -adic group then follow immediately for the transform of  $G$  from the selfsame properties for  $G$ —hence the  $m$ -adic group  $G'$ . In general, two  $m$ -adic groups  $G$  and  $G'$  may be said to be *simply isomorphic* if a 1-1 correspondence can be set up between their elements such that if  $s'$  of  $G'$  is the correspondent of  $s$  in  $G$ , then we will have, for all elements of  $G$ ,

$$[c(s_1 s_2 \cdots s_m)]' = c'(s'_1 s'_2 \cdots s'_m),$$

$c$  and  $c'$  designating the  $m$ -adic operations of  $G$  and  $G'$  respectively. For  $G'$  the transform of  $G$  this is immediate with  $c$  and  $c'$  the common unexpressed  $m$ -adic operation.

We reserve a more detailed treatment of transforms for our study of finite polyadic groups, and turn to the question of invariance. An  $m$ -adic element, polyad, or group will be said to be invariant under an  $i$ -ad if it is transformed into itself by that  $i$ -ad. It will then be said to be invariant under an  $m$ -adic group if it is invariant under every polyad of that group. Since  $G^*$  is generated by  $G$ , it follows that for  $K$  to be invariant under  $G$ , it is sufficient that it be invariant under every element of  $G$ . If such a  $K$  is an element (subgroup) of  $G$  it will then be said to be an invariant element (subgroup) of  $G$ . Clearly, the condition that an  $m$ -group  $G$  be abelian is equivalent to each of its elements being an invariant element of  $G$ . For, in the notation of the coset theorem,  $\{s_1, s_2\}$  and  $\{s_2, s_1\}$  being equivalent becomes  $s_1 s_2 = s_2 s_1$ , or,  $s_2^{-1} s_1 s_2 = s_1$ ; and conversely.

Given an invariant subgroup  $H$  of  $G$ , the expansion of  $G$  in cosets as regards  $H$  immediately leads to an  $m$ -adic quotient group  $G/H$ . In fact, since  $H$  is invariant under  $G$ , it immediately follows that  $H_0$ , the associated 2-group of  $H$ , is also invariant under  $G$ ; that is,  $H_0$ , as subgroup of  $G^*$ , is invariant under each element of  $G$  considered as element of  $G^*$ . For  $H_0$  consists of all products of  $m-1$  elements chosen arbitrarily and independently from  $H$ . Hence the transform of  $H_0$  under any element  $s$  of  $G$  consists of all products of  $m-1$  elements chosen arbitrarily and independently from the transform of  $H$  under  $s$ , that is, from  $H$  all over again.

Consider then the expansion in cosets  $G = H_0 s_\alpha + H_0 s_\beta + \cdots + H_0 s_\epsilon$ . Then, exactly as in ordinary group theory, the coset in which the element  $s_1 s_2 \cdots s_m$  appears depends only on the cosets containing the elements  $s_1, s_2, \cdots, s_m$ . If then  $\sigma_1, \sigma_2, \cdots, \sigma_m$  represent the cosets containing  $s_1, s_2, \cdots, s_m$  respectively, we may write the coset containing  $s_1 s_2 \cdots s_m$  in the form  $\sigma_1 \sigma_2 \cdots \sigma_m$ . An  $m$ -adic operation is thus determined on these cosets as elements; and, again as in classic theory, these cosets constitute an  $m$ -adic group under this operation. We may therefore call this group the quotient group  $G/H$ .

As we shall see later,  $m$ -adic quotient groups arising from invariant subgroups are very special kinds of polyadic groups. However, Dörnte has em-

phasized that  $m$ -adic quotient groups can arise in more general fashion. In our presentation, his argument reduces to the fact that the only use made of the invariance of subgroup  $H$  under  $G$  was to prove the invariance of  $H_0$  under  $G$ . We shall call a subgroup  $H$  of  $G$  whose associated 2-group  $H_0$  is invariant under  $G$  a *semi-invariant* subgroup of  $G$ . It follows that every semi-invariant subgroup of an  $m$ -adic group leads to an  $m$ -adic quotient group.

This result can be made still more general. For we observed earlier that any subgroup  $H_0$  of the associated 2-group  $G_0$  of  $G$  gives rise to expansions in cosets. It therefore follows that *every subgroup of the associated 2-group of an  $m$ -adic group which is invariant under the  $m$ -adic group leads to an  $m$ -adic quotient group*. In the absence of a subgroup  $H$  of  $G$  we shall write this quotient group  $G/H_0$ .

It is immediately seen that with  $H_0$  thus invariant under  $G$ , the right cosets of  $G$  as regards  $H_0$  are identical with the left cosets. For  $s^{-1}H_0s = H_0$  yields  $H_0s = sH_0$ . Conversely, if the right cosets of  $G$  as regards  $H_0$  are identical with the left cosets, then, for each element  $s$  of  $G$ ,  $H_0s = sH_0$ , so that  $H_0$  is invariant under  $G$ . We thus see that the Dörnte concept of semi-invariance may be said to be the necessary and sufficient condition that a subgroup of a polyadic group give rise to a quotient group. Our extension, however, frees  $G$  from the need of possessing a subgroup  $H$  corresponding to the  $H_0$  invariant under  $G$ .

In recent literature the concept of homomorphism appears as essentially equivalent to that of quotient group<sup>(25)</sup>. By means of our coset theorem we readily show the same to be true for  $m$ -groups<sup>(26)</sup>. As the analysis is not too immediate, we have refrained from explicitly using this concept except in the last section where it is especially needed.

An  $m$ -group  $G$  with operation  $c$  may be said to be *homomorphic* to an  $m$ -group  $\bar{G}$  with operation  $\bar{c}$  if there is a many-one correspondence between the elements of  $G$  and of  $\bar{G}$  such that whenever  $s_1, s_2, \dots, s_m$  of  $G$  respectively correspond to  $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_m$  of  $\bar{G}$ ,  $c(s_1s_2 \dots s_m)$  corresponds to  $\bar{c}(\bar{s}_1\bar{s}_2 \dots \bar{s}_m)$ . We first show that such a homomorphism between  $G$  and  $\bar{G}$  determines a homomorphism between their abstract containing groups  $G^*$  and  $\bar{G}^*$ . In fact, let  $i$ -ad  $r$  of  $G^*$  be said to correspond to  $i$ -ad  $\bar{r}$  of  $\bar{G}^*$  if there exist elements  $s_1, s_2, \dots, s_i$  of  $G$ , and corresponding elements  $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_i$  of  $\bar{G}$ , such that  $r = c^*(s_1s_2 \dots s_i)$ ,  $\bar{r} = \bar{c}^*(\bar{s}_1\bar{s}_2 \dots \bar{s}_i)$ . It is readily seen that this sets up a correspondence between all the elements of  $G^*$  and all the elements of  $\bar{G}^*$ . Furthermore, this correspondence is many-one. For suppose  $r$  of  $G^*$  corresponds to  $\bar{r}_1$  and  $\bar{r}_2$  of  $\bar{G}^*$ . Then we must have  $r = c^*(s_1s_2 \dots s_i)$ ,  $\bar{r}_1 = \bar{c}^*(\bar{s}_1\bar{s}_2 \dots \bar{s}_i)$ , and, also,  $r = c^*(s'_1s'_2 \dots s'_i)$ ,  $\bar{r}_2 = \bar{c}^*(\bar{s}'_1\bar{s}'_2 \dots \bar{s}'_i)$ , with  $s_1, s_2, \dots, s_i, s'_1, s'_2, \dots, s'_i$  of  $G$  corresponding to  $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_i, \bar{s}'_1, \bar{s}'_2, \dots, \bar{s}'_i$  respectively of  $\bar{G}$ . If then  $s$  of  $G$  corresponds to  $\bar{s}$  of  $\bar{G}$ , the equation

<sup>(25)</sup> See, for example, B. L. van der Waerden, *Moderne Algebra*, Berlin, 1930, vol. 1, §9.

<sup>(26)</sup> Dörnte's Theorem 8, §6, does the same for his more limited concept of  $m$ -adic quotient group under the assumption that the homomorph has at least one "first order element."

$c(s_1 s_2 \cdots s_i s \cdots s) = c(s'_1 s'_2 \cdots s'_i s \cdots s)$ , obtained from the two forms of  $r$ , yields  $\bar{c}(\bar{s}_1 \bar{s}_2 \cdots \bar{s}_i \bar{s} \cdots \bar{s}) = \bar{c}(\bar{s}'_1 \bar{s}'_2 \cdots \bar{s}'_i \bar{s} \cdots \bar{s})$  as a result of the homomorphism between  $G$  and  $\bar{G}$ . Hence  $\bar{r}_1 = \bar{r}_2$ . Finally, if  $r_1$  and  $r_2$  of  $G^*$  thus correspond to  $\bar{r}_1$  and  $\bar{r}_2$  of  $\bar{G}^*$ ,  $c^*(r_1 r_2)$  corresponds to  $\bar{c}^*(\bar{r}_1 \bar{r}_2)$ —immediately, if  $r_1$  and  $r_2$  are an  $i$ -ad and  $j$ -ad respectively with  $i+j \leq m-1$ , and via the homomorphism between  $G$  and  $\bar{G}$  if  $i+j > m-1$ . The many-one correspondence between the elements of  $G^*$  and of  $\bar{G}^*$  is therefore a homomorphism.

The ordinary theorem on homomorphisms is therefore applicable, and we can state that the elements of  $G^*$  corresponding to the identity of  $\bar{G}^*$  constitute an invariant subgroup  $H_0$  of  $G^*$ , while the elements of  $G^*$  corresponding to any element of  $\bar{G}^*$  constitute a coset in the expansion of  $G^*$  as regards  $H_0$ , the quotient group  $G^*/H_0$  being then simply isomorphic with  $\bar{G}^*$ . Since the identity of  $\bar{G}^*$  is an  $(m-1)$ -ad,  $H_0$  must consist of  $(m-1)$ -ads in  $G^*$ , and is thus a subgroup of  $G_0$  invariant under  $G$ . Those cosets of  $G^*$  as regards  $H_0$  which involve elements of  $G$  therefore constitute an expansion of  $G$  as regards  $H_0$ . Finally, the correspondence between  $G^*$  and  $\bar{G}^*$  is but the original correspondence for elements of  $G$  and  $\bar{G}$ . We thus have the following theorem. *If  $m$ -group  $G$  is homomorphic to  $m$ -group  $\bar{G}$ , there is an  $m$ -adic quotient group  $G/H_0$  such that the correspondents of each element of  $\bar{G}$  constitute a coset in  $G/H_0$ , this quotient group then being simply isomorphic with  $\bar{G}$ .* Actually, as we have seen,  $H_0$  consists of the elements of  $G_0$  corresponding to the identity of  $\bar{G}_0$  in the homomorphism between  $G^*$  and  $\bar{G}^*$ , and hence between  $G_0$  and  $\bar{G}_0$  determined by the given homomorphism. Since an  $m$ -group  $G$  is clearly homomorphic to any  $m$ -adic quotient group  $G/H_0$ , the equivalence of the concepts of homomorphism and quotient group has been shown to hold also for  $m$ -groups.

A homomorphism between  $m$ -groups  $G$  and  $\bar{G}$  is thus always an  $(N, 1)$  isomorphism with fixed  $N$ ,  $N$  of course finite for finite  $m$ -groups. A more immediate consequence of the given homomorphism is that it sets up a many-one correspondence between the subgroups of  $G$  and the subgroups of  $\bar{G}$ , an  $m$ -group being considered now as a subgroup of itself. In fact, given a subgroup of  $G$ , the corresponding elements of  $\bar{G}$  are readily seen to satisfy the conditions for an  $m$ -group, and thus constitute the uniquely corresponding subgroup of  $\bar{G}$ . On the other hand, given a subgroup of  $\bar{G}$ , the set of all corresponding elements of  $G$  constitutes a subgroup of  $G$  with the given subgroup of  $\bar{G}$  as corresponding subgroup, and indeed, contains all such subgroups of  $G$ . Clearly this many-one correspondence between the subgroups of  $G$  and of  $\bar{G}$  is preserved under the relation "subgroup of"—subgroup, in the above sense of group or subgroup.

It is also readily verified that if the set  $\bar{G}$  is not known to be an  $m$ -group under operation  $\bar{c}$ , yet the remainder of the definition of homomorphism between  $G$  and  $\bar{G}$  is satisfied, then  $\bar{G}$  is an  $m$ -group under  $\bar{c}$ , and hence the given relation a genuine homomorphism. In fact, the only part of our defi-

dition of  $m$ -group not immediately given for  $\bar{G}$  under  $\bar{c}$ , as a consequence of its being satisfied by  $G$  under  $c$ , is the uniqueness of the solution of  $\bar{c}(\bar{s}_1 \bar{s}_2 \cdots \bar{s}_m) = \bar{s}_{m+1}$  for  $\bar{s}_i$ ,  $1 \leq i \leq m$ . Passing by the considerations of the footnote of §1 and a special argument valid only for  $\bar{G}$  finite, we can in every case solve corresponding equations  $c(s_1 s_2 \cdots s_m) = s_{m+1}$  for  $s_i$  as in §2, with all  $s$ 's except  $s_i$  and  $s_{m+1}$  fixed, and thus find that all such  $s$ 's must correspond to the same, consequently unique,  $\bar{s}_i$  <sup>(27)</sup>.

Our converse of the coset theorem admits of immediate extension to the case of an  $m$ -adic quotient group. For the statement of this result we need the concept of order, when finite, of an element of an  $m$ -group as given in the beginning of §21. We may note now, however, that an element  $s$  may be said to be of first order if  $c(ss \cdots s) = s$ , the unit class with sole member  $s$  then being a subgroup of the given  $m$ -group. We see then immediately that *if an element of an  $m$ -adic quotient group is of the first order, the corresponding coset constitutes a subgroup of the given  $m$ -group*. For the isomorphism between the given  $m$ -group and the quotient group shows that if in an equation  $c(s_1 s_2 \cdots s_m) = s_{m+1}$  any  $m$  elements are in the coset, the  $(m+1)$ -st element must also be in that coset. Now consider any element  $\sigma$  of finite order  $k$  of the quotient group. Anticipating a concept of the next section, we may note now that our given  $m$ -group will constitute a polyadic group under the extended operation  $c(s_1 s_2 \cdots s_\mu)$  with  $\mu = k(m-1) + 1$ . Our  $m$ -adic quotient group likewise extends to a  $\mu$ -group with the element  $\sigma$  now being a first order element of the  $\mu$ -adic quotient group. The previous result therefore leads to the following. *If an element of an  $m$ -adic quotient group is of finite order  $k$ , then the elements of the corresponding coset constitute a polyadic group under the operation of the given group extended to  $k(m-1) + 1$  elements.*

**5. Reducibility.** Given any ordinary group with class of elements  $C$  and dyadic operation  $s_1 s_2$ , an  $m$ -adic group on the same elements will be determined if we set up the  $m$ -adic operation  $c(s_1 s_2 \cdots s_m) = s_1 s_2 \cdots s_m$ . We shall

<sup>(27)</sup> If a general isomorphism between  $m$ -groups  $G$  and  $\bar{G}$  be defined as a many-many correspondence between their elements in which  $m$ -adic products of corresponding elements correspond, then, for finite  $m$ -adic groups, as for finite ordinary groups, the correspondence is that of a simple isomorphism between  $m$ -adic quotient groups of  $G$  and  $\bar{G}$ . On the other hand, Dickson (these Transactions, vol. 6 (1905), pp. 205-208) has shown by an example that the finite group theorem does not hold for infinite groups, while Loewy (Festschrift Heinrich Weber, 1912, pp. 198-227) calls an isomorphism "vollständig" if inverses of corresponding elements also correspond—the case when the finite group theorem does hold for infinite groups—and derives a number of interesting conditions for a general isomorphism to be "vollständig." In the case of infinite  $m$ -adic groups, the condition under which the finite  $m$ -adic group theorem goes over can be written in a variety of ways, but perhaps most symmetrically as follows. If in two equations  $c(s_1 s_2 \cdots s_m) = s_{m+1}$ ,  $\bar{c}(s'_1 s'_2 \cdots s'_m) = s'_{m+1}$ ,  $m$  of the  $m+1$  symbols in the first equation, and the  $m$  corresponding symbols in the second equation, represent elements of  $G$  and  $\bar{G}$  respectively that correspond, then the elements represented by the remaining symbols must correspond. The writer is indebted to Reinhold Baer for the above references (as well as for the Neumann reference of §30).



call the  $m$ -group an extension of the 2-group, and say that it is reducible to that 2-group. Note that while the coset theorem presented an arbitrary polyadic group in a somewhat similar light, the elements of the polyadic group formed but a proper subclass of the class of elements of the 2-group; whereas, when a polyadic group is reducible to a 2-group, the classes of elements are identical.

More generally, given a  $\mu$ -group with class of elements  $C$  and operation  $c_\mu(s_1 s_2 \cdots s_\mu)$ , if  $m$  is any number in the form  $k(\mu-1)+1$  we can form the extended operation  $c_\mu(s_1 s_2 \cdots s_m) = c_\mu(s_1 s_2 \cdots s_{\mu-1} c_\mu(s_\mu s_{\mu+1} \cdots s_{2\mu-2} (\cdots c_\mu(s_{(k-1)(\mu-1)+1} s_{(k-1)(\mu-1)+2} \cdots s_{k(\mu-1)+1} \cdots)))$ . The members of  $C$  will then form an  $m$ -adic group under the operation  $c_m(s_1 s_2 \cdots s_m) = c_\mu(s_1 s_2 \cdots s_m)$ . As before, the  $m$ -group will be said to be an extension of the  $\mu$ -group, and reducible to the  $\mu$ -group.

An  $m$ -adic operation on a finite number of elements is most naturally exhibited by an  $m$ -dimensional table. We shall therefore say that an  $m$ -adic group is of dimension  $m$ . We then see that while a 2-group has an extension for each dimension  $m > 2$ , a  $\mu$ -group has an extension for those and only those dimensions  $m$  for which  $m-1$  is a multiple of  $\mu-1$ .

A given  $m$ -group will be said to be *reducible to a  $\mu$ -group* if there exists a  $\mu$ -group to which it is reducible. The  $m$ -group will be said to be irreducible if it is not reducible to a  $\mu$ -group for any  $\mu < m$ <sup>(28)</sup>. Dörnte has already given a necessary and sufficient condition that a polyadic group be reducible to a 2-group. We proceed to generalize this result to reducibility to a  $\mu$ -group.

A  $(\mu-1)$ -ad  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  will be said to be commutative with an element  $a$  if the  $\mu$ -ads  $\{a_1, a_2, \cdots, a_{\mu-1}, a\}$  and  $\{a, a_1, a_2, \cdots, a_{\mu-1}\}$  are equivalent. We then have the following basic theorem on reducibility. *A necessary and sufficient condition that a given  $m$ -group be reducible to a  $\mu$ -group,  $m = k(\mu-1)+1$ , is that there be a  $(\mu-1)$ -ad  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  formed from elements of the  $m$ -group such that the  $(\mu-1)$ -ad is commutative with every element of the  $m$ -group, and such that the  $(m-1)$ -ad  $\{a_1, a_2, \cdots, a_{\mu-1}, a_1, a_2, \cdots, a_{\mu-1}, \cdots, a_1, a_2, \cdots, a_{\mu-1}\}$  is an identity of the  $m$ -group.*

The necessity of this condition follows immediately from the existence and properties of identities. For, if the  $m$ -group is reducible to a  $\mu$ -group, let  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  be an identity of such a  $\mu$ -group. If  $c_\mu$  is the operation of the  $\mu$ -group,  $c_\mu(a_1 a_2 \cdots a_{\mu-1} s) = s = c_\mu(s a_1 a_2 \cdots a_{\mu-1})$  for every element  $s$  of the  $\mu$ -group. Hence  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  is commutative with every element of the  $\mu$ -group, and hence, by the hypothesis of reducibility, with every element of the  $m$ -group. Furthermore, the  $(m-1)$ -ad  $\{a_1, a_2, \cdots, a_{\mu-1}, a_1, a_2, \cdots, a_{\mu-1}, \cdots, a_1, a_2, \cdots, a_{\mu-1}\}$  is an extended identity of the  $\mu$ -group, and hence an identity of the  $m$ -group, as was to be proved.

As for the sufficiency of the condition, with  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  as in the

<sup>(28)</sup> "Echt" in Dörnte. Otherwise, "unecht" or "ableitbar."



hypothesis, define the  $\mu$ -adic operation

$$c_\mu(s_1 s_2 \cdots s_\mu) = c_m(s_1 s_2 \cdots s_\mu a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}).$$

We proceed to prove that the elements of the  $m$ -group constitute a  $\mu$ -group under the operation  $c_\mu$ , and that the given  $m$ -group is reducible to this  $\mu$ -group. Of the two conditions defining a polyadic group, condition 1 is satisfied by the proposed  $\mu$ -group as an immediate consequence of its being satisfied by the given  $m$ -group. On the other hand, condition 2 for the  $\mu$ -group becomes

$$\begin{aligned} c_m(c_m(s_1 s_2 \cdots s_\mu a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}) s_{\mu+1} \\ \cdots s_{2\mu-1} a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}) \\ = c_m(s_1 c_m(s_2 s_3 \cdots s_{\mu+1} a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}) s_{\mu+2} \\ \cdots s_{2\mu-1} a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}) \\ \cdots \\ = c_m(s_1 s_2 \cdots s_{\mu-1} c_m(s_\mu s_{\mu+1} \cdots s_{2\mu-1} a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}) a_1 a_2 \\ \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}), \end{aligned}$$

which follows from condition 2 for the  $m$ -group, and the commutativity of  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  with each element of the  $m$ -group. Hence the existence of the  $\mu$ -group. Finally, using extended operations, and applying the commutativity part of our hypothesis, we will have

$$c_\mu(s_1 s_2 \cdots s_m) = c_m(s_1 s_2 \cdots s_m a_1 a_2 \cdots a_{\mu-1} \cdots a_1 a_2 \cdots a_{\mu-1}) = c_m(s_1 s_2 \cdots s_m),$$

the second expression involving a sequence consisting of  $k(k-1)$  sequences  $a_1 a_2 \cdots a_{\mu-1}$ , which sequence, therefore, constitutes an extended identity of the  $m$ -group—since by hypothesis  $k$  such sequences constitute an identity. Hence the reducibility of the  $m$ -group to the  $\mu$ -group follows.

From the definition of  $c_\mu$ , we see that the  $(\mu-1)$ -ad  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  is indeed an identity of the resulting  $\mu$ -group.

The above theorem may be used to prove a polyadic group irreducible, as is shown by the following simple illustration. The class of integers constitutes an infinite  $m$ -adic group under the operation  $s_1 + s_2 + \cdots + s_m + 1$ . Since the group is abelian, reducibility to a  $\mu$ -group with  $m = k(\mu-1) + 1$  is equivalent to the existence of an integer  $a$  such that  $ka + s + 1 = s$ , that is,  $ka = -1$ , which is impossible for any integral  $k > 1$ . Hence the  $m$ -group is irreducible.

The commutativity condition can be restated to read  $\{a_1, a_2, \cdots, a_{\mu-1}\}$  is invariant under the  $m$ -group. Since the present multiplicity of basic operations makes us refrain from employing the simplifications of the coset theorem, the concept of invariance is preferable only for  $\mu-1=1$ . Our  $(\mu-1)$ -ad is now a single element  $a$ ; and the further condition that the  $(m-1)$ -ad

$\{a, a, \dots, a\}$  be an identity of the  $m$ -group may be restated to read:  $a$  is of first order. For this condition is equivalent to  $c_m(aa \dots aa) = a$ . We may therefore state the special result, a rewording only of Dörnte's, *a necessary and sufficient condition that a given  $m$ -group be reducible to an ordinary group is that the  $m$ -group possess an invariant element of first order*. Our succeeding development will reveal many general classes of polyadic groups that can be proved reducible to 2-groups. One such class is already at hand, that is, *all  $m$ -adic quotient groups arising from invariant subgroups of  $m$ -adic groups are reducible to 2-groups*. For the element of the quotient group corresponding to the invariant subgroup is immediately seen to be invariant under the quotient group, and of  $m$ -adic order one. In this connection we may observe that semi-invariant subgroups also lead to special kinds of polyadic quotient groups, for the element corresponding to that semi-invariant subgroup must again be of first order. On the other hand, any polyadic group can be a quotient group in our most general sense; for, with  $H_0$  the identity of  $G_0$ ,  $G/H_0$  is identical with  $G$ .

Given an  $m$ -adic group  $G$ , we may ask for the distribution of, and interrelations between, the polyadic groups to which it is reducible. Note immediately that if  $G$  is reducible to  $G'$ , and  $G'$  to  $G''$ ,  $G$  is reducible to  $G''$ , so that the class of groups to which  $G'$  is reducible is a subclass of the class of groups to which  $G$  is reducible whenever  $G$  is reducible to  $G'$ . Our results are of two kinds, both derived from the above theorem.

The first type of result is not much more than a restatement of the condition of the theorem. We recall that, if  $G$  is reducible to  $G'$ , the class of elements of  $G$  is identical with the class of elements of  $G'$ , while the operation of  $G$  is an extended operation of  $G'$ . It follows that a class of equivalent  $i$ -ads of  $G$  is also a class of equivalent  $i$ -ads of  $G'$ , and conversely. In particular, the class of identities of  $G'$  is a class of equivalent polyads<sup>(29)</sup> of  $G$ , so that the classes of identities of two groups to which  $G$  may be reducible are either the same or mutually exclusive.

When the classes of identities are distinct, the two groups in question will be distinct, as their operations cannot then be identical<sup>(30)</sup>. On the other hand, we easily see that when the classes of identities are the same, the groups are identical. For, if their operations are  $c'$  and  $c''$ , then, with  $\{a_1, a_2, \dots, a_{m-1}\}$  an identity of each, we have

$$c'(s_1 s_2 \dots s_m) = c(s_1 s_2 \dots s_m a_1 a_2 \dots a_{m-1} \dots a_1 a_2 \dots a_{m-1}) = c''(s_1 s_2 \dots s_m),$$

<sup>(29)</sup> By a class of equivalent polyads we mean a class of equivalent  $i$ -ads for some fixed  $i$ . While the elements of  $G^*$  as first written are classes of equivalent  $i$ -ads with  $1 \leq i \leq m-1$ , in general no such restriction is intended by the above phrase. As suggested in §2, by the use of extended operations the concept of equivalent  $i$ -ads becomes valid for  $i > m-1$ . This observation will be of greater importance later in the present section.

<sup>(30)</sup> They may however be "abstractly the same" in the sense of being simply isomorphic. See the opening paragraph of §23.

$c$  being an extended operation of each group. Observe finally that in the sufficiency proof of our basic theorem, and in the succeeding observation, if  $\{a_1, a_2, \dots, a_{\mu-1}\}$  satisfies the given condition of that theorem, each  $(\mu-1)$ -ad equivalent to  $\{a_1, a_2, \dots, a_{\mu-1}\}$  also does. We therefore can state the following result. *There is a 1-1 correspondence between the groups to which a given  $m$ -adic group is reducible and the classes of equivalent polyads satisfying the condition of the basic theorem, each such class of equivalent polyads being the class of identities of the corresponding group.*

In particular, there are as many 2-groups to which an  $m$ -adic group is reducible as there are invariant elements of order one in the  $m$ -group<sup>(31)</sup>. Thus, consider an ordinary abelian group of finite order  $g$ . If  $d$  is any divisor of  $g$ , there are at least  $d$  elements  $a$  in this 2-group with  $a^d = 1$ . If this 2-group be extended to a  $(d+1)$ -group, each such element  $a$  is of order one in the  $(d+1)$ -group, and invariant therein. The  $(d+1)$ -group is therefore reducible to at least  $d$  distinct 2-groups, each such  $a$ , in fact, being the identity of the corresponding 2-group.

Our second type of result concerns the possible dimensions of the groups to which a given polyadic group is reducible. The complete result is an immediate consequence of the following theorem. *If an  $m$ -group is reducible to a  $\mu_1$ -group and a  $\mu_2$ -group, it is reducible to a  $\mu$ -group where  $\mu-1$  is the highest common factor of  $\mu_1-1$  and  $\mu_2-1$ .* To prove this theorem let  $\{a'_1, a'_2, \dots, a'_{\mu_1-1}\}$  and  $\{a''_1, a''_2, \dots, a''_{\mu_2-1}\}$  be identities of the  $\mu_1$ -group and  $\mu_2$ -group respectively. They then satisfy the condition of our basic theorem. Furthermore, all but one of the letters in each can be chosen arbitrarily.

If then  $\mu_1 > \mu_2$ , we may assume  $a'_1 = a''_1, \dots, a'_{\mu_2-1} = a''_{\mu_2-1}$ . Consider then the sequence  $\{a'_{\mu_2}, \dots, a'_{\mu_1-1}\}$  which we shall write  $\{a'''_1, \dots, a'''_{\mu_1-\mu_2}\}$ , with  $\mu_3-1 = (\mu_1-1) - (\mu_2-1)$ . Then all but one of the letters of this sequence are arbitrary. Inductively, we thus obtain the sequence  $\{a^{(\lambda)}_1, \dots, a^{(\lambda)}_{\mu_\lambda-1}\}$ , with all but one letter arbitrary, from the sequence  $\{a^{(\lambda-1)}_1, \dots, a^{(\lambda-1)}_{\mu_{\lambda-1}-1}\}$  and the smallest preceding sequence, easily seen to be unique. Clearly the process terminates when and only when  $\mu_\lambda-1$  is equal to the smallest preceding  $\mu$ .

Now in terms of the  $\mu_\lambda-1$ 's, this process is nothing more than the Euclid algorithm for finding the highest common factor of  $\mu_1-1$  and  $\mu_2-1$ , where the process of division is replaced by the more primitive form of repeated subtractions. Hence, the above process terminates, and the last sequence found may be written  $\{a_1, \dots, a_{\mu-1}\}$ , where  $\mu-1$  is the highest common factor of  $\mu_1-1$  and  $\mu_2-1$ . We now prove that such a  $(\mu-1)$ -ad satisfies the condition of our basic theorem.

First, the sequence  $\{a'''_1, \dots, a'''_{\mu_3-1}\}$  is commutative with every element of the given  $m$ -group. For we have  $\{a'_1, \dots, a'_{\mu_1-1}\} = \{a''_1, \dots, a''_{\mu_2-1}, a'''_1, \dots, a'''_{\mu_3-1}\}$ , so that  $c(a'''_1 \dots a'''_{\mu_3-1} a'''_{\mu_2-1} \dots a'''_{\mu_1-1} s_1 s_2 \dots s_m) = c(s_1 a'''_1 \dots$

<sup>(31)</sup> In the case of abelian triadic groups this reduces to a theorem of Lehmer's.

$a_{\mu_2-1}'' a_1''' \cdots a_{\mu_2-1}''' s_{l+1} \cdots s_m) = c(a_1'' \cdots a_{\mu_2-1}'' s_l a_1''' \cdots a_{\mu_2-1}''' s_{l+1} \cdots s_m)$ . Hence, by induction, each  $\{a_1^{(\lambda)}, \cdots, a_{\mu_\lambda-1}^{(\lambda)}\}$  is commutative with every element of the  $m$ -group, and so  $\{a_1, \cdots, a_{\mu-1}\}$  also is thus commutative.

As for the second part of the condition, clearly  $m-1 = k(\mu-1)$  with integral  $k$ . As in the commutativity argument, and with the commutativity property, we obtain from the extended identities consisting of  $k$  sequences  $\{a_1', \cdots, a_{\mu-1}'\}$  and  $k$  sequences  $\{a_1'', \cdots, a_{\mu-1}''\}$  an extended identity consisting of  $k$  sequences  $\{a_1', \cdots, a_{\mu-1}'\}$ . By induction,  $k$  sequences  $\{a_1^{(\lambda)}, \cdots, a_{\mu-1}^{(\lambda)}\}$  constitute an extended identity for every  $\lambda$ , and hence the same is true of  $k$  sequences  $\{a_1, \cdots, a_{\mu-1}\}$ . But, since  $k(\mu-1) = m-1$ , the last is indeed an identity of our given  $m$ -group.  $\{a_1, \cdots, a_{\mu-1}\}$  therefore satisfies completely the condition of our basic theorem, whence the present result.

It follows that if  $\mu_0$  is the least dimension of the groups to which a given  $m$ -group is reducible, all other dimensions  $\mu$  of such groups must be such that  $\mu-1$  is a multiple of  $\mu_0-1$ . We shall call  $\mu_0$  the *real dimension* of the  $m$ -group, with, of course,  $\mu_0 = m$  if the group is irreducible. Since every  $\mu-1$  must also be a divisor of  $m-1$ , we easily obtain the following solution of the problem of the distribution of the dimensions of the groups to which a given polyadic group is reducible. *If a group of dimension  $m$  has real dimension  $\mu_0$ , and we write  $m-1 = k_0(\mu_0-1)$ , then the dimensions of the groups to which the  $m$ -group is reducible are those and only those numbers  $\mu$  for which  $\mu-1 = k(\mu_0-1)$ ,  $k$  a proper divisor of  $k_0$ .*

While this result justifies the term *real dimension* on the basis of a mere enumeration of distinct dimensions, other considerations show that an  $m$ -group in general, even if reducible, must still be considered an  $m$ -group. We have already given an example which shows that the same  $m$ -group may be reducible to different groups of the same dimension, and, indeed, of the real dimension of the  $m$ -group. We now further observe that an  $m$ -group may be reducible to an irreducible group of higher dimension than the real dimension of the  $m$ -group, that is, not every succession of reductions of a group need lead to the real dimension of the group. If we call the dimensions of the irreducible groups to which a polyadic group is reducible the *irreducible dimensions* of the given group, the real dimension of the group is only the smallest of its irreducible dimensions.

In contrast with the class of groups to which an  $m$ -group is reducible, the class of extensions of an  $m$ -group is of very simple structure, since it has one and only one group of each dimension  $\mu$  with  $\mu-1$  a multiple of  $m-1$ , and no others. Of course, the reason is that extension is the direct process, reduction indirect. We now combine these processes to yield the concept of derived group.

Given an  $m$ -group  $G$ , a polyadic group  $G'$  will be said to be *derivable from  $G$*  if it can be obtained from  $G$  by a finite succession of extensions and reductions.

The class of all polyadic groups derivable from a given polyadic group will be called a *net* of polyadic groups. From this definition we see that each group of a net yields that net. Furthermore, all groups of a given net have the same class of elements; only the operations differ.

The concept of a net of polyadic groups is considerably simplified by the following result. *Any group of a net can be obtained from any other by a single extension followed by a single reduction.* A single extension or a single reduction can obviously be replaced by an extension followed by a reduction. Since two successive extensions are equivalent to a single extension, two successive reductions to a single reduction, our result will follow if we can show that a reduction followed by an extension is equivalent to an extension followed by a reduction. Let then  $G'$  with operation  $c'_m$  be reducible to  $G''$  with operation  $c''_m$ , and let  $G''$  be extended to  $G'''$  with operation  $c'''_m$ . With the above subscripts designating dimensionality, we have  $m' - 1 = k'(m'' - 1)$ ,  $m''' - 1 = k''(m'' - 1)$ . Now  $c'_m$  and  $c'''_m$  are both extensions of operation  $c''_m$ . If then we extend  $c'''_m$  to an operation  $c^{IV}_m$  with  $m^{IV} - 1 = k'k''(m'' - 1)$ ,  $c^{IV}_m$  will be an extension of both  $c'_m$  and  $c''_m$ . The corresponding group  $G^{IV}$  is then reducible to both  $G'$  and  $G''$ , whence our result.

Stated otherwise, *given any two groups of a net there is a third group of the net reducible to each of the given groups.* We could therefore redefine a net as the class of groups to which the extensions of a given group are reducible, though the conclusion that a net does not depend on the particular group in it chosen as the given group is then not immediate.

The two types of results referred to in the case of the groups to which a given group is reducible now easily lead to corresponding results for the net of groups derivable from a given group. In this connection, a  $(\mu - 1)$ -ad  $\{a_1, a_2, \dots, a_{\mu-1}\}$  of an  $m$ -group will be said to be of finite order if some polyad of the form  $\{a_1, a_2, \dots, a_{\mu-1}, a_1, a_2, \dots, a_{\mu-1}, \dots, a_1, a_2, \dots, a_{\mu-1}\}$  is an extended identity of the  $m$ -group. We then easily prove the following. *There is a 1-1 correspondence between the groups of the net of groups derivable from a given group and the classes of equivalent polyads of finite order which are commutative with every element of the given group, each such class of equivalent polyads then being the class of identities of the corresponding group<sup>(22)</sup>.* In fact, the above redefinition of a net immediately yields a many-one correspondence of the above type, which is then seen to be one-one due to any pair of groups of a net being in the class of groups to which a third is reducible.

Actually, it is easily verified that each of the concepts: class of equivalent polyads, commutative with every element, and even polyad of finite order, is independent of the particular group of the net chosen as given group, so that the above result can be restated in terms of the net alone. It is also easily proved that for finite polyadic groups every polyad is of finite order, so that

(22) Here, as elsewhere, "group" unqualified means polyadic group.



in such cases the corresponding condition need not be explicitly stated. In particular, there are as many 2-groups in the net as there are invariant elements of finite order, and hence, for finite polyadic groups, as many as there are invariant elements.

We pause to prove explicitly that the transform of one element of a group of a net by another is independent of the particular group employed. This will be so if true of any pair of groups, one reducible to the other. Since the operation of one of these groups is an extended operation of the other, an identity of the first group is an extended identity of the second; hence an inverse of an element in the first, an extended inverse of that element in the second, whence the identical transforms.

The second type of result is obtained still more easily. We shall call the least dimension of the groups of a net their *outer real dimension*. The outer real dimension of a group is then always less than or equal to its real dimension. Given an  $m$ -group  $G$  of outer real dimension  $\mu^0$ , some third group  $G'$  of the net will be reducible both to the  $m$ -group, and a group of dimension  $\mu^0$ . The real dimension of  $G'$  will therefore exactly equal  $\mu^0$ . As  $G'$  is reducible to  $G$ , we see that  $m-1$  is a multiple of  $\mu^0-1$ . That is, if the outer real dimension of an  $m$ -group is  $\mu^0$ , then  $\mu^0-1$  must be a divisor of  $m-1$ .

Hence, also, all the groups of the net have dimensions  $\mu$  with  $\mu-1$  a multiple of  $\mu^0-1$ . Since, from a group of dimension  $\mu^0$ , mere extensions yield groups of all such dimensions, we have the following main result. *If the outer real dimension of the groups of a net is  $\mu^0$ , their dimensions are those and only those numbers  $\mu$  for which  $\mu-1=k(\mu^0-1)$ .*

The first type of result is easily restated to yield a criterion for determining the outer real dimension of a group. In particular, *the outer real dimension of a group is 2 when and only when it contains an invariant element of finite order*. Thus, a finite abelian polyadic group is always of outer real dimension 2, and so is derivable from a 2-group, while a group having no invariant element is always of outer real dimension greater than 2. The existence of the latter type of group is peculiar to polyadic theory. A simple example is furnished by the class of odd substitutions of the symmetric group of degree three. By the converse of the coset theorem they form a triadic group of order three under the product of three substitutions as operation, and yet involve no invariant element. The three elements, incidentally, are all of first order in the triadic group.

As in the case of mere reducibility, we shall call the dimensions of the irreducible groups of a net the *outer irreducible dimensions* of each group in the net. By contrast, a dimension will be said to be a *reducible dimension* of the groups of the net if there is at least one group of the net of that dimension, while all such groups are reducible. While we have no general theorem giving the distribution of these dimensions, the following special results lend a certain insight into the possibilities involved.



First, a group may have its real dimension as its only outer irreducible dimension. This is readily proved to be so for any 2-group which has no invariant element other than the identity. In this case, in fact, the net of groups consists only of the 2-group, and its extensions.

By contrast, a group may have an infinite number of outer irreducible dimensions. Thus it can be shown that for the ordinary cyclic group of order two the outer irreducible dimensions are the infinite set of numbers of the form  $2^n + 1$ ,  $n = 0, 1, 2, \dots$ .

Finally, it can be shown that every finite polyadic group has an infinite number of reducible dimensions. To be specific, if an  $m$ -group has  $g$  elements, there is, of course, at least one group of the net of dimension  $(kg+1)(m-1)+1$ , for each  $k=1, 2, 3, \dots$ , and every group of the net of such a dimension is reducible, reducible to dimension  $m$ , in fact.

We append a brief discussion of the generalization of the concept of a net of groups that arises from a consideration of the subgroups of a group. Let the *complex* of groups obtainable from a given polyadic group be the class of all polyadic groups obtainable from the given group by finite successions of the three operations "extension of," "reduction of," and "subgroup of." It is readily verified by means of the very concepts involved that an extension of a subgroup of a group is also a subgroup of an extension of a group; and that a subgroup of a reduction of a group is also a reduction of a subgroup of the group. It follows that *any group in a complex can be obtained from the given group by an operation of the single form "extension of" followed by "subgroup of" followed by "reduction of" if not merely by "extension of" followed by "reduction of."*

In the case of abelian groups we further have that a reduction of a subgroup of a group is also a subgroup of a reduction of the group, a result obtainable with the help of our criterion of reducibility. It follows that *the complex of groups obtainable from an abelian polyadic group consists of the groups in the corresponding net of groups, and their subgroups*. That this is not true for all complexes can be seen from the case of a group with a first order element, but no invariant element. For the first order element constitutes a subgroup of the given group reducible to a 2-group; while, the outer real dimension of the given group being greater than 2, the dimensions of all the groups in the net, and hence of their subgroups, is greater than 2.

It is readily seen that the groups of a complex whose classes of elements are the same as that of the original group constitute the net of that group, or, as we shall now phrase it, the net of the complex. Clearly the net of a complex also consists of all of its groups from which that complex is obtainable. On the other hand, a group of a complex with class of elements a proper subclass of that of the original group will yield a complex which is a proper subclass of the given complex, and may be called a subcomplex thereof. If we call the nets of the subcomplexes of a complex the subnets of that complex,

then it is clear that the net and subnets of a complex constitute a separation of the groups of the complex into mutually exclusive sets.

The relationship between the subcomplexes of a complex is in part furnished by the following result. *If of two groups in a complex the class of elements of the first group is contained in the class of elements of the second, then the first group is in the complex obtained from the second.* For consider the two groups to be obtained from an initial group according to our first result. Using  $(c_m, C)$  to designate a group with  $m$ -adic operation  $c_m$  and class of elements  $C$ , we may indicate the process as follows:

$$\begin{aligned}(c_m, C) &\rightarrow (c'_m, C) \rightarrow (c'_m, C') \rightarrow (c''_m, C'), \\(c_m, C) &\rightarrow (c''''_m, C) \rightarrow (c''''_m, C'') \rightarrow (c^{IV}_m, C'').\end{aligned}$$

The two groups in the second column are also reductions of a third group  $(c^{IV}_m, C)$ . Since the third column symbolizes groups, it follows that  $(c^{IV}_m, C')$  and  $(c^{IV}_m, C'')$  are groups; and as  $C'$  is contained in  $C''$  by hypothesis,  $(c^{IV}_m, C')$  is a subgroup of  $(c^{IV}_m, C'')$ , if not identical with it. Now  $(c^{IV}_m, C')$ ,  $(c'_m, C')$  and  $(c''''_m, C')$  are in a single net of groups, as are also  $(c^{IV}_m, C'')$ ,  $(c'_m, C'')$  and  $(c''''_m, C'')$ . Hence  $(c''_m, C')$  is in the complex obtainable from  $(c^{IV}_m, C'')$ , as was to be proved.

A particular application of the above result is the following. *Any two groups of a complex which have the same class of elements are derivable from each other, that is, belong to one and the same net.* It follows that there is a 1-1 correspondence between the subnets, including the net, into which the groups of a complex were separated, and the different classes of elements of the groups in the complex.

Hence also, or directly from our general result, there is a 1-1 correspondence between the subcomplexes, including the complex, of a complex, and the different classes of elements of the groups in the complex, each complex being obtainable from those and only those groups whose classes of elements are identical with the class of elements corresponding to the complex. Moreover, our general result shows that one subcomplex contains a second when and only when the class of elements corresponding to the first contains the class of elements corresponding to the second. We now complete this picture by proving the following. *If two subcomplexes  $K'$  and  $K''$  of a complex correspond to the classes of elements  $C'$  and  $C''$ , then the logical product of  $K'$  and  $K''$ , null when the logical product of  $C'$  and  $C''$  is null, is otherwise a complex, namely the complex corresponding to the logical product of  $C'$  and  $C''$ .* For  $C'$  and  $C''$  must be the classes of elements of two groups  $(c^{IV}_m, C')$  and  $(c^{IV}_m, C'')$  of the complex. In the notation of the previous proof,  $(c^{IV}_m, C')$  and  $(c^{IV}_m, C'')$  are then groups of the complex. If then  $C'''$ , the logical product of  $C'$  and  $C''$ , is not null,  $(c^{IV}_m, C''')$  is a group of the complex. The case  $C'''$  null is immediate. Otherwise, then, there will be a subcomplex  $K'''$  corresponding to  $C'''$ .

Our earlier result then shows immediately that a group  $G$  is common to  $K'$  and  $K''$  when and only when it is in  $K'''$ .

Further results on the subcomplexes of a complex obtained from a finite polyadic group, and more particularly a finite abelian polyadic group, will be found at the end of §22, our second section on cyclic polyadic groups<sup>(32)</sup>.

**6. Arbitrary containing ordinary groups.** The coset theorem led to the abstract containing ordinary group  $G^*$  of an  $m$ -group  $G$  merely by a consideration of  $G$  treated abstractly. Often, however, the elements of  $G$  may immediately be given in such a form that the  $m$ -adic operation is but an extension of a more primitive dyadic operation, as when  $G$  is an  $m$ -adic group of ordinary substitutions. In such a case a containing 2-group arises directly, and may be more useful than the abstract containing group.

A 2-group  $G^*$  will be called a *containing group* of an  $m$ -group  $G$  if the elements of  $G$  are among the elements of  $G^*$ , the operation of  $G$  an extension of the operation of  $G^*$ , while  $G^*$  is generated by the elements of  $G$ . In what follows we simultaneously investigate the possible structure of  $G^*$ , and its relationship to  $G$ . We must therefore explicitly distinguish between their operations  $c^*$  and  $c^*$  respectively<sup>(34)</sup>.

Let two polyads  $\{s_1, s_2, \dots, s_i\}$  and  $\{s'_1, s'_2, \dots, s'_i\}$  of  $G$  lead to identical products in  $G^*$ ; that is, let  $c^*(s_1 s_2 \dots s_i) = c^*(s'_1 s'_2 \dots s'_i)$ . Since  $i' - i$  must then be a multiple of  $m - 1$ , we can annex elements  $s''_1, \dots, s''_{j'}$  of  $G$ , if need be, so that the resulting equation  $c^*(s_1 s_2 \dots s_i s''_1 \dots s''_{j'}) = c^*(s'_1 s'_2 \dots s'_i s''_1 \dots s''_{j'})$  can be rewritten  $c(s_1 s_2 \dots s_i s''_1 \dots s''_{j'}) = c(s'_1 s'_2 \dots s'_i s''_1 \dots s''_{j'})$  in, perhaps, extended notation. But this equation can now be written  $c^*(s_1 s_2 \dots s_i s''_1 \dots s''_{j'}) = c^*(s'_1 s'_2 \dots s'_i s''_1 \dots s''_{j'})$ , whence we obtain  $c^*(s_1 s_2 \dots s_i) = c^*(s'_1 s'_2 \dots s'_i)$ . That is, if two polyads of  $G$  lead to identical products in  $G^*$  they lead to identical products in  $G^*$ . If then we let every element of the form  $c^*(s_1 s_2 \dots s_i)$  in  $G^*$  correspond to element  $c^*(s_1 s_2 \dots s_i)$  of  $G^*$ , a one-many correspondence is set up between those elements of  $G^*$  and of  $G^*$  which are obtainable as products of elements of  $G$ .

This correspondence is clearly preserved under the respective operations of these groups. For if  $r_1$  and  $r_2$  of  $G^*$  correspond to  $r'_1$  and  $r'_2$  respectively

<sup>(32)</sup> The development of the section just ended, lengthy as it is, is probably but one of many possible developments leading to sets of related polyadic groups. Dörnte's Theorem 7, §2, can probably be made the starting point for such a different development. The possibilities are further widened if a theory is contemplated which would include the relationship between a polyadic group and the corresponding "schar."

<sup>(34)</sup> It might be thought that now, when the ordinary group demanded by Miller's theorem is immediately given, at least the structure of  $G^*$  requires no further investigation. But, apart from the fact that Miller's theorem is given for finite groups, his hypothesis that for some integer  $n$  the products of any  $n$  but no fewer elements of  $G$  is in  $G$  is not immediately given, but is replaced by  $G$ 's being an  $m$ -group. As we also need the relationship between  $G^*$  and  $G^*$ , we make our development entirely independent of Miller's.

of  $G^{**}$ , by writing these elements as corresponding products of elements in  $G$  we see immediately that  $c^*(r_1 r_2)$  corresponds to  $c^{**}(r'_1 r'_2)$ . Since  $G^*$  consists of the products of elements in  $G$ , it easily follows that the products in  $G^{**}$  of elements of  $G$  themselves constitute a group which can then be none other than  $G^{**}$ ; for  $G^{**}$  is generated by  $G$ . Furthermore our one-many correspondence, which is therefore a correspondence between all the elements of  $G^{**}$  and of  $G^*$ , is indeed a one-many isomorphism between  $G^{**}$  and  $G^*$ .

For fixed  $i$  we shall call the set of elements of  $G^{**}$  which are the products of  $i$  elements of  $G$  the  $i$ th coset of  $G^{**}$ . For these elements the above set of equations can be reversed so that our one-many correspondence between  $G^{**}$  and  $G^*$  becomes a 1-1 correspondence between the elements of the  $i$ th cosets of  $G^{**}$  and of  $G^*$  for each  $i \geq 1$ . From the corresponding result for  $G^*$ , it follows that the elements of the  $i$ th coset of  $G^{**}$  will be obtained in 1-1 fashion if in the expression  $c^{**}(s_1 \cdots s_{i-1}s)$  we let  $s_1, \dots, s_{i-1}$  be arbitrary fixed elements of  $G$ , and let  $s$  run through  $G$ .

Let now  $k$  designate the least  $i$  for which the corresponding coset of  $G^{**}$  contains the identity  $I'$  of  $G^{**}$ . It follows, first, that the first  $k$  cosets of  $G^{**}$  are mutually exclusive. For if we could have  $c^{**}(s_1 \cdots s_i) = c^{**}(s'_1 \cdots s'_j)$  with  $1 \leq i < j \leq k$ , then, by rewriting  $c^{**}(s'_1 \cdots s'_j)$  in the form  $c^{**}(s_1 \cdots s_i s'_{i+1} \cdots s'_j)$ , we would have  $c^{**}(s'_{i+1} \cdots s'_j) = I'$ , in contradiction to our definition of  $k$ . On the other hand, the  $(k+1)$ -st coset of  $G^{**}$  is identical with the first, that is, with  $G$ , for we can write its elements in the form  $c^{**}(s_1 \cdots s_k s)$  with  $c^{**}(s_1 \cdots s_k) = I'$ . Hence also the  $(k+2)$ -nd coset is identical with the 2d, and so on.  $G^{**}$  therefore consists of the elements of its first  $k$  cosets, while succeeding cosets are cyclic repetitions of these. In particular, the  $(m-1)$ -st coset must be identical with the  $k$ th coset. For if  $\{s_1, s_2, \dots, s_{m-1}\}$  is an identity of  $G$ ,  $c^{**}(s_1 s_2 \cdots s_{m-1}) = I'$ , so that the  $(m-1)$ -st and  $k$ th cosets have an element in common. Hence  $k$  is a divisor of  $m-1$ .

Returning to our correspondence between the elements of  $G^{**}$  and of  $G^*$  we see that it is 1-1 between the elements of  $G^{**}$  and the elements of the first  $k$  cosets of  $G^*$ , and of each succeeding set of  $k$  cosets of  $G^*$ . Our one-many correspondence is thus actually  $[1, (m-1)/k]$ , and we therefore have a  $[1, (m-1)/k]$  isomorphism between  $G^{**}$  and  $G^*$ . To complete our analysis we consider the analogue in  $G^{**}$  of the associated 2-group  $G_0$  of  $G$  in  $G^*$ .

Our  $[1, (m-1)/k]$  correspondence is clearly 1-1 between the elements of the  $k$ th coset of  $G^{**}$ , and of  $G_0$ , the  $(m-1)$ -st coset of  $G^*$ . Since the product of two elements of the  $k$ th coset of  $G^{**}$  is in the  $2k$ th coset, and hence also in the  $k$ th coset, of  $G^{**}$ , the previous  $[1, (m-1)/k]$  isomorphism between  $G^{**}$  and  $G^*$  is simple between the  $k$ th coset of  $G^{**}$ , and  $G_0$ . It follows that the  $k$ th coset of  $G^{**}$  constitutes a group with operation  $c^{**}$  simply isomorphic with  $G_0$ . We shall call it the associated ordinary group of  $G$  in  $G^{**}$ , and symbolize it  $G'_0$ . The same argument used in proving  $G_0$  invariant under  $G^*$  shows  $G'_0$  to be invariant under  $G^{**}$ .

Since the  $i$ th coset of  $G^{**}$  is given by  $c^{**}(s_1 \cdots s_{i-1}s)$ , with  $s_1, \dots, s_{i-1}$  fixed elements of  $G$ ,  $s$  running through  $G$ , we can let  $s_1, \dots, s_{i-1}$  be the same element  $s_0$  of  $G$ , and write that  $i$ th coset  $s_0^{i-1}G$  in ordinary notation. It can likewise be written  $Gs_0^{i-1}$ . We thus obtain the expansion  $G^{**} = G + Gs_0 + Gs_0^2 + \cdots + Gs_0^{k-1}$ . Since  $Gs_0^{k-1} = G'_0$ , and  $Gs_0^k = G$ , we therefore have

$$G = G'_0 s_0,$$

while the above expansion becomes

$$G^{**} = G'_0 s_0 + G'_0 s_0^2 + \cdots + G'_0 s_0^{k-1} + G'_0.$$

But this is the expansion of  $G^{**}$  in augmented cosets as regards the invariant subgroup  $G'_0$ , assuming  $G'_0$  is not itself  $G^{**}$ . It follows that the quotient group  $G^{**}/G'_0$  is of index  $k$ , while the element in that quotient group corresponding to  $G$  generates  $G^{**}/G'_0$ .

This concludes our discussion of the structure of  $G^{**}$ . As for its isomorphism with  $G^*$ , observe first that in that isomorphism elements of  $G$  correspond to themselves. We then see that the isomorphism between  $G^{**}$  and  $G^*$  is determined by this partial correspondence provided  $k$ , and the element of the  $k$ th coset of  $G^{**}$  which serves as the identity of  $G^{**}$ , are specified. For the correspondence between elements of  $G$  and themselves determines the 1-1 correspondence between the elements of the  $i$ th cosets of  $G^{**}$  and of  $G^*$  for every  $i$ . And given  $k$ , and  $c^{**}(s_1^0 s_2^0 \cdots s_k^0) = I'$ ,  $s$ 's in  $G$ , if  $j = \kappa k + l$ ,  $1 \leq l \leq k$ , the equation  $c^{**}(s_1^0 s_2^0 \cdots s_k^0 \cdots s_1^0 s_2^0 \cdots s_k^0 s_1 s_2 \cdots s_l) = c^{**}(s_1 s_2 \cdots s_l)$  serves to identify each symbolized element of the  $j$ th coset of  $G^{**}$  with a unique element of the  $l$ th coset, and thus completes the correspondence between the elements of  $G^{**}$  and  $G^*$ . In particular, the simple isomorphism between  $G'_0$  and  $G_0$  is also thus determined. We therefore have the following comprehensive theorem:

*Every containing 2-group  $G^{**}$  of an  $m$ -group  $G$ , if not itself a 2-group  $G'_0$  to which  $G$  is reducible, contains an invariant subgroup  $G'_0$  of index  $k$ , with  $k$  a divisor of  $m-1$ ,  $G$  a coset of  $G^{**}$  as regards  $G'_0$ , and the quotient group  $G^{**}/G'_0$  generated by the element corresponding to  $G$ . Furthermore,  $G^{**}$  admits a  $[1, (m-1)/k]$  isomorphism with  $G^*$ , the abstract containing 2-group of  $G$ , which reduces to a simple isomorphism between  $G'_0$  and  $G_0$ , the associated 2-group of  $G$ . This isomorphism makes each element of  $G$  correspond to itself, and is, in fact, determined by this correspondence when  $k$ , which is the smallest  $i$  for which an  $i$ -ad of  $G$  yields the identity of  $G^{**}$ , as well as the class of equivalent  $k$ -ads of  $G$  thus yielding the identity of  $G^{**}$ , are specified.*

We shall call  $k$  the *index* of the containing 2-group. We have then, in particular, that any two containing groups of index  $m-1$  of an  $m$ -group are simply isomorphic, the isomorphism in question making each element of the  $m$ -group correspond to itself, and being in turn determined by this correspondence. Hence,



any containing group of index  $m-1$  of an  $m$ -group  $G$  may be considered to be the abstract containing group  $G^*$  of  $G$ .

We further have that *any two containing groups of index 1 of an  $m$ -group are simply isomorphic*. For the  $G^{*'}s$  are then also the  $G'_0s$  which are both simply isomorphic with  $G_0$ . Observe, however, that the simple isomorphism now no longer makes elements of  $G$  correspond to themselves, or the  $G^{*'}s$  would be identical. In fact, a different element of  $G$  serves as identity in each  $G^{*'}$ . Since  $G$  is now reducible to  $G^{*'}$ , and conversely, we have as a corollary the following result on the 2-groups to which an  $m$ -group is reducible, and hence also on the 2-groups in a net. *All 2-groups in a net of groups are simply isomorphic*.

Before considering the same question for two containing groups of index  $k$ ,  $1 < k < m-1$ , we ask when an  $m$ -group will admit a containing 2-group of index  $k$ . We then easily obtain the following theorem. *A necessary and sufficient condition that an  $m$ -group admit a containing group of index  $k$ ,  $k < m-1$ , is that the  $m$ -group be reducible to a  $(k+1)$ -group*. In fact, the observation that in a containing group  $G^{*'}$  of index  $k$  the products of  $k+1$  elements of  $G$  must be in  $G$  is easily extended to show that the elements of  $G$  constitute a  $(k+1)$ -group under the operation  $c^{*'}(s_1 s_2 \cdots s_{k+1})$ . As  $k$  is a divisor of  $m-1$ , the operation  $c(s_1 s_2 \cdots s_m) = c^{*'}(s_1 s_2 \cdots s_m)$  is an extension of  $c^{*'}(s_1 s_2 \cdots s_{k+1})$ , and, consequently,  $G$  is reducible to the corresponding  $(k+1)$ -group. Conversely, if  $G$  is reducible to a  $(k+1)$ -group, the abstract containing group of the  $(k+1)$ -group is of index  $k$ . But this group is clearly also a containing group of  $G$ , and of index  $k$ . In particular, *an irreducible  $m$ -group admits containing groups of index  $m-1$  only, and conversely*. Hence, the abstract containing group of an irreducible polyadic group may be said to be its only containing group.

This relation to reducibility shows that there are as many essentially different containing groups of index  $k < m-1$  of an  $m$ -group  $G$  as there are  $(k+1)$ -groups to which  $G$  is reducible. Hence when  $1 < k < m-1$ , as when  $k=1$ , two essentially different containing groups of index  $k$  will not admit a simple isomorphism which makes each element of  $G$  correspond to itself, since the classes of equivalent  $k$ -ads yielding their identities will be different. Moreover, unlike the case  $k=1$ , they need not even admit a simple isomorphism which transforms the class of elements of  $G$  into itself. For our example of a group having an infinite number of outer irreducible dimensions easily leads to a group  $G$  reducible to two groups  $G_1$  and  $G_2$  of the same dimension, one reducible, the other irreducible. The abstract containing groups of  $G_1$  and  $G_2$  are containing groups of  $G$  of the same index; and did they admit a simple isomorphism of the type in question,  $G_1$  and  $G_2$  would be simply isomorphic, and hence could not be one reducible, the other irreducible.

Finally, a word about the application of arbitrary containing groups of an  $m$ -group to the study of the  $m$ -group. With the containing group  $G^{*'}$  specified,





and all the displacements of the letters in the equations defining the semi-abelianism.

Observe immediately that for  $m=2$  there is no semi-abelianism distinct from abelianism. For in some equation a pair of letters  $s_i, s_j$  will appear in different orders on opposite sides of the equation; and by replacing all other letters by the identity we obtain the condition for abelianism  $s_i s_j = s_j s_i$ . This serves to make plausible our general result, and to give a hint of its proof.

In the general case, then, let  $G$  be any  $m$ -group semi-abelian according to a given formal type, and let some letter  $s_j$  have a nonzero displacement  $k$  in one of the equations defining that semi-abelianism. Since re-symbolization allows either member of the equation to be written first, we may write the equation

$$s_1 \cdots s_{j-1} s_j s_{j+1} \cdots s_l = s_{i_1} \cdots s_{i_{j+k-1}} s_j s_{i_{j+k}} s_{i_{j+k+1}} \cdots s_{i_l},$$

so that

$$s_j = [(s_1 \cdots s_{j-1})^{-1} s_{i_1} \cdots s_{i_{j+k-1}}] s_j [s_{i_{j+k}} \cdots s_{i_l} (s_{j+1} \cdots s_l)^{-1}].$$

The first bracket is equivalent to some  $k$ -ad  $s' s'' \cdots s^{(k)}$ . Since at least one letter inside that bracket and outside the parenthesis must be different from all the letters in the parenthesis, that  $k$ -ad, and hence  $s', s'', \dots, s^{(k)}$ , can be arbitrary. The second bracket is equivalent to some  $\kappa$ -ad  $\bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa)}$ . We can always assume  $\kappa > 1$ , by introducing an identity if need be, and hence at least  $\bar{s}^{(\kappa)}$  is arbitrary. That is, for every  $s', s'', \dots, s^{(k)}, \bar{s}^{(\kappa)}$ , we can find  $\bar{s}', \bar{s}'', \dots, \bar{s}^{(\kappa-1)}$  so that

$$s_j = s' s'' \cdots s^{(k)} s_j \bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)}$$

for every  $s_j$ . Letting  $s_j = s'$ , we find that  $s'' \cdots s^{(k)} s' \bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)} = 1$ , whence

$$\bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} \bar{s}^{(\kappa)} s'' \cdots s^{(k)} s' = 1.$$

Letting  $s_j = \bar{s}^{(\kappa)}$ , we find  $s' s'' \cdots s^{(k)} \bar{s}^{(\kappa)} \bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} = 1$ , whence

$$\bar{s}' \bar{s}'' \cdots \bar{s}^{(\kappa-1)} s' s'' \cdots s^{(k)} \bar{s}^{(\kappa)} = 1.$$

It follows that for every  $s', s'', \dots, s^{(k)}, \bar{s}^{(\kappa)}$  in  $G$ ,

$$s' s'' \cdots s^{(k)} \bar{s}^{(\kappa)} = \bar{s}^{(\kappa)} s' s'' \cdots s^{(k)} s'.$$

Dropping momentarily the condition  $\mu - 1$  a divisor of  $m - 1$  in our definition of  $\mu$ -semi-abelianism, we have therefore proved that for each displacement  $k > 0$ ,  $G$  is  $(k+1)$ -semi-abelian.

Let now  $G$  be  $(k_1+1)$ -semi-abelian and  $(k_2+1)$ -semi-abelian. We then prove that  $G$  is  $(k+1)$ -semi-abelian with  $k = \text{H.C.F.}(k_1, k_2)$ . This will follow if for every such  $k_1$  and  $k_2$  with  $k_2 > k_1$ ,  $G$  is  $(k_3+1)$ -semi-abelian with  $k_3 = k_2 - k_1$ . But under our hypothesis, with all other letters unmoved, we have

$s_1 \cdots s_{k+1} \cdots s_{k+1} = s_{k+1} \cdots s_{k+1} \cdots s_1 = s_{k+1} \cdots s_{k+1} \cdots s_1 = s_1 \cdots s_{k+1} \cdots s_{k+1}$ . Hence  $s_{k+1} \cdots s_{k+1} = s_{k+1} \cdots s_{k+1}$  as desired.

Finally, we show that if the  $m$ -group  $G$  is  $(k+1)$ -semi-abelian, it is also  $(k'+1)$ -semi-abelian with  $k' = \text{H.C.F.}(k, m-1)$ . Since  $G$  is  $(k+1)$ -semi-abelian, it is also  $(\kappa k+1)$ -semi-abelian for every positive integral  $\kappa$ . It is therefore also  $(k''+1)$ -semi-abelian with  $k''$  any positive integer in the form  $\kappa k - \lambda(m-1)$ . For in the equation defining the  $(\kappa k+1)$ -semi-abelianism there are at least  $\lambda(m-1)$  letters between the first and last letters of each member; and by choosing  $\lambda(m-1)$  of these letters consecutively to form an extended identity the desired  $(k''+1)$ -semi-abelianism is revealed. As positive integers  $\kappa$  and  $\lambda$  can always be chosen so that  $\kappa k - \lambda(m-1) = \text{H.C.F.}(k, m-1)$ , our result follows.

From these three special results it follows that every  $m$ -group possessing a given formal type of semi-abelianism is  $\mu$ -semi-abelian with  $\mu$  as in the statement of our theorem. It remains to be shown that every  $m$ -group that is  $\mu$ -semi-abelian also satisfies the given formal semi-abelianism. For each of the given equations separates the letters in the left side of the equation into  $\mu-1$  mutually exclusive sets such that each set consists of all letters whose "distance" from a given letter is a multiple of  $\mu-1$ . Since in passing from the left side to the right side of the equation each letter suffers a displacement itself a multiple of  $\mu-1$ , the result is to permute the letters of each set among themselves. Now a single application of our hypothesis of  $\mu$ -semi-abelianism to the left side of the equation in question constitutes a transposition of two letters in the same set. As  $\mu$ -semi-abelianism implies  $[\kappa(\mu-1)+1]$ -semi-abelianism, every such transposition can be effected. And, as any substitution is the product of transpositions, successive applications of our hypothesis of semi-abelianism will transform the left side of each equation so that each of its  $\mu-1$  sets assumes the form it has on the right. That is, each equation of the given formal semi-abelianism will be satisfied by the elements of any  $m$ -group that is  $\mu$ -semi-abelian. The equivalence in question has therefore been demonstrated.

That  $\mu$ -semi-abelianism is a different type of semi-abelianism for different divisors  $\mu-1$  of  $m-1$  is readily proved by examples. By the theorem of the next section, an  $m$ -group  $G = G_0 s_0$  will be determined by the following hypothesis:  $G_0$  an ordinary cyclic group of order  $2^{m-1}-1$  generated by  $t$ ,  $s_0^{m-1} = 1$ ,  $s_0^{-1} t s_0 = t^2$ . Since  $G_0$  is abelian, the first result of the next paragraph shows  $G$  to be  $m$ -semi-abelian. Now a similar argument shows an  $m$ -group  $G$  to be  $\mu$ -semi-abelian,  $\mu-1$  a divisor of  $m-1$ , when and only when the  $(\mu-1)$ -ads of  $G$  are commutative with the  $(m-1)$ -ads of  $G$ . Since  $s_0^{m-1}$  is the first ordinary positive power of  $s_0$  commutative with  $t$ , it follows that  $G$  is not  $\mu$ -semi-abelian for any divisor  $\mu-1$  of  $m-1$  other than  $m-1$ . Now let  $\mu_1-1$ ,  $\mu_2-1$  be any two distinct divisors of  $m-1$  with, say,  $\mu_1 > \mu_2$ . By the preceding method construct a  $\mu_1$ -group  $G'$  which is  $\mu_1$ -semi-abelian, but not  $\mu_2$ -semi-

abelian for any divisor  $\mu_3 - 1$  of  $\mu_1 - 1$  other than  $\mu_1 - 1$ . The extension of  $G'$  to an  $m$ -group  $G''$  then has the same property. It then follows that the  $m$ -group  $G''$  while  $\mu_1$ -semi-abelian is not  $\mu_2$ -semi-abelian, since otherwise it would be  $\mu_3$ -semi-abelian with  $\mu_3 - 1 = \text{H.C.F.}(\mu_1 - 1, \mu_2 - 1)$ , and thus a divisor of  $\mu_1 - 1$  other than  $\mu_1 - 1$ . The  $m$ -group  $G''$  thus shows  $\mu_1$ -semi-abelianism to be not equivalent to  $\mu_2$ -semi-abelianism whenever  $\mu_1 \neq \mu_2$ . Coupled with our previous theorem it yields the following result. *There are as many distinct types of semi-abelianism for  $m$ -adic groups as there are distinct divisors of  $m - 1$ .*

In what follows we restrict our attention to ordinary, that is,  $m$ -semi-abelianism, a property implied by any type of semi-abelianism. Since the associated ordinary group  $G_0$  of an  $m$ -group  $G$  consists of the products of  $m - 1$  arbitrary elements of  $G$ , the condition that  $G_0$  is abelian is a condition of semi-abelianism on  $G$  of formal type

$$s_1 s_2 \cdots s_{m-1} s_m s_{m+1} \cdots s_{2m-2} = s_m s_{m+1} \cdots s_{2m-2} s_1 s_2 \cdots s_{m-1}.$$

As each letter suffers a displacement  $m - 1$ , by our general result this type of semi-abelianism is equivalent to  $m$ -semi-abelianism. Hence, *every semi-abelian  $m$ -group has an abelian associated group, and conversely*. If an element  $s$  of a semi-abelian group  $G$  is invariant under  $G$ , it is also invariant under  $G_0$ , and hence  $G = G_0 s$  is abelian. That is, *if a semi-abelian  $m$ -group is non-abelian, it has no invariant element*. If  $s_1$  and  $s_2$  are any two elements of semi-abelian  $G$ ,  $t$  any element of  $G_0$ , then, since  $s_1 = t^{-1} s_2$ , with  $t'$  in  $G_0$ , and since  $t$  and  $t'$  are commutative, we have  $s_1^{-1} t s_1 = s_2^{-1} t s_2$ . Hence, *all the elements of a semi-abelian  $m$ -group  $G$  transform an arbitrary given element of the associated group  $G_0$  into the same element*. Now let  $H$  be any subgroup of semi-abelian  $G$ . Its associated subgroup  $H_0$  is then invariant under any element  $s_0$  of  $H$ . But every element  $s$  of  $G$  transforms the elements of  $H_0$  as does  $s_0$ . Hence  $H_0$  is invariant under  $G$ . That is, *every subgroup of a semi-abelian group is semi-invariant*<sup>(25)</sup>.

**8. On the construction of polyadic groups.** We proceed to prove the following general theorem on the construction of abstract polyadic groups referred to in connection with the converse of the coset theorem. *Given any abstract 2-group  $G_0$  to serve as associated group, an abstract element  $s_0$  subject to the condition  $s_0^{m-1} = t_0$ ,  $t_0$  in  $G_0$ , and any automorphism  $T$  of  $G_0$ , which carries  $t_0$  into itself, and whose  $(m - 1)$ -st power is the automorphism of  $G_0$  under  $t_0$ , to serve as the automorphism of  $G_0$  under  $s_0$ , then there is one and only one corresponding abstract  $m$ -group  $G$ ; conversely every  $m$ -group can be thus determined*<sup>(26)</sup>.

<sup>(25)</sup> See Dörnte's §7 for quite a different set of properties of semi-abelian groups. Dörnte's result that a triadic group consisting of first order elements only must be semi-abelian is equivalent for finite groups to a result of Miller's as a consequence of the above equivalence of the semi-abelianism of  $G$ , and abelianism of  $G_0$ . By introducing the polyadic groups  $G_i$  of our §34 to take the place of  $G_0$  in the discussion of the last paragraph, the results of that paragraph can be specifically generalized to  $\mu$ -semi-abelianism.

<sup>(26)</sup> After this theorem was obtained by the writer, a closely related result was published by Turing as an illustration of a more general theorem in the theory of group extensions. (Not

For the second part of this theorem note that given an  $m$ -group  $G$ , and any  $s_0$  in  $G$ ,  $G_0$ ,  $t_0$ , and  $T$  are determined, and obviously satisfy the conditions of the theorem. It follows from the first part of the succeeding proof that  $G$  is determinable as stated.

We turn then to the first part of the theorem. For purposes of analysis, consider the coset representation of a hypothetical  $G$  satisfying the given conditions. We would then have  $G = G_0 s_0$ . If we write the elements of  $G_0$  as  $t_i$ , we may correspondingly symbolize the elements of  $G$  by  $s_i$ , with  $s_i = t_i s_0$ . Of course  $s_0$  must then be identified with that  $s_i$  for which  $t_i$  is the identity of  $G_0$ , while  $t_0$  will appear as some  $t_k$ . We must then have, for the operation of  $G$ ,

$$\begin{aligned} c(s_{i_1} s_{i_2} \cdots s_{i_m}) &= t_{i_1} s_0 t_{i_2} s_0 \cdots t_{i_m} s_0 = t_{i_1} (s_0 t_{i_2} s_0^{-1}) \cdots (s_0^{m-1} t_{i_m} s_0^{-m+1}) s_0 \\ &= (t_{i_1} \cdot T^{-1} t_{i_2} \cdots T^{-(m-1)} t_{i_m} \cdot t_0) s_0, \end{aligned}$$

so that  $c(s_{i_1} s_{i_2} \cdots s_{i_m})$ , and with it  $G$ , if it exists, is completely determined by our hypothesis.

We next prove that the elements  $s_i = t_i s_0$  actually constitute an  $m$ -group under this operation. As to condition 1 of the definition of an  $m$ -group, given  $c(s_{i_1} s_{i_2} \cdots s_{i_m}) = s_{i_{m+1}}$  with all  $s$ 's but  $s_{ij}$  specified members of  $G$ , we correspondingly have  $t_{i_1} \cdot T^{-1} t_{i_2} \cdots T^{-(m-1)} t_{i_m} \cdot t_0 = t_{i_{m+1}}$ , with all elements specified members of  $G_0$  with the exception of  $t_{i_{m+1}}$ , when  $j = m+1$ ,  $T^{-(j-1)} t_{ij}$ , when  $j \neq m+1$ . In the first case, a unique  $t_{i_{m+1}}$  in  $G_0$ , and, hence  $s_{i_{m+1}}$  in  $G$ , are immediately determined. In the second case, a unique  $T^{-(j-1)} t_{ij}$  in  $G_0$  is determined, hence again  $t_{ij}$  in  $G_0$ , and  $s_{ij}$  in  $G$ . As for condition 2, we have

$$\begin{aligned} c(s_{i_1} \cdots s_{i_{j-1}} c(s_{i_j} \cdots s_{i_{j+m-1}}) s_{i_{j+m}} \cdots s_{i_{2m-1}}) \\ = (t_{i_1} \cdots T^{-(j-2)} t_{i_{j-1}} \cdot T^{-(j-1)} (t_{i_j} \cdots T^{-(m-1)} t_{i_{j+m-1}} \cdot t_0) \\ \quad \cdot T^{-j} t_{i_{j+m}} \cdots T^{-(m-1)} t_{i_{2m-1}} \cdot t_0) s_0 \\ = (t_{i_1} \cdots T^{-(j-2)} t_{i_{j-1}} \cdot T^{-(j-1)} t_{i_j} \cdots T^{-(j+m-2)} t_{i_{j+m-1}} \\ \quad \cdot T^{-(j+m-1)} t_{i_{j+m}} \cdots T^{-(2m-2)} t_{i_{2m-1}} \cdot t_0^2) s_0, \end{aligned}$$

the last since  $T^{-(j-1)} t_0 = t_0$ , and  $t_0 \cdot t = T^{-(m-1)} t \cdot t_0$ , by our hypothesis. The result is thus independent of  $j$ , whence follows condition 2.

It remains to be shown that the  $m$ -group  $G$  thus obtained actually re-determines, via  $s_0$ , the  $G_0$ ,  $t_0$ ,  $T$  of the given hypothesis<sup>(27)</sup>. From the operation  $c$

to be confused with our polyadic concept of §5. See A. M. Turing, *The extensions of a group*, *Compositio Mathematica*, vol. 5 (1938), pp. 357-367.) From this point of view, the abstract containing groups of  $m$ -groups with given  $G_0$  are the extensions of  $G_0$  by the cyclic group of order  $m-1$ . Our theorem on the determination of  $G$  could then have been based on the determination of  $G^*$  as cyclic extension of  $G_0$ . The theorem on cyclic extensions thus envisaged would be not quite Turing's (Theorem 5, loc. cit.), but equivalent thereto by the identification of our  $T$  with his  $\xi$ ,  $t_0$  with  $\delta^{-1} \tau^*$ .

<sup>(27)</sup> In connection with the preceding footnote it must be mentioned that this part of the proof was overlooked by the writer until the final check-up on the entire paper.



as given, and again with the aid of the relation  $T^{-(m-1)t_{i_m}} \cdot t_0 = t_0 \cdot t_{i_m}$ , we see that equivalent  $(m-1)$ -ads  $\{s_{i_1}, s_{i_2}, \dots, s_{i_{m-1}}\}$  are those for which the corresponding elements  $t_{i_1} \cdot T^{-1}t_{i_2} \cdot \dots \cdot T^{-(m-2)}t_{i_{m-1}} \cdot t_0$  of the given 2-group  $G_0$  are the same. If then we represent the elements of the associated 2-group of  $G$  thus by the elements of the given group  $G_0$ , and determine the operation of this associated group via  $[\{s_{i_1}, s_{i_2}, \dots, s_{i_{m-1}}\}] \cdot [\{s_{j_1}, s_{j_2}, \dots, s_{j_{m-1}}\}] = [\{c(s_{i_1}s_{i_2} \dots s_{i_{m-1}}s_{j_1}), s_{j_2}, \dots, s_{j_{m-1}}\}]$ , bracket meaning class of  $(m-1)$ -ads equivalent to the specified  $(m-1)$ -ad, we find this operation, again with the help of the above relation, to be identical with the operation of the given 2-group. That is, abstractly, the given  $G_0$  is the associated ordinary group of  $G$ . Since, for the  $s_i = s_0$ ,  $t_i$  is the identity, we immediately have  $s_0^{m-1} = [\{s_0, s_0, \dots, s_0\}] = t_0$  in the above representation. Finally, by introducing identity, hence inverse, and thus transform, in their original polyadic form, it can likewise be shown that if the elements of  $G_0$  are transformed by  $s_0$  the resulting automorphism of  $G_0$  is  $T$ . Therefore, the proof has been completed.

We have already used the converse of the coset theorem in giving an example of a 3-group of order three having no variant element. This 3-group can now be given abstractly in accordance with the above theorem. For  $G_0$ , take the cyclic group  $(1, t, t^2)$ . Let  $s_0^2 = 1$ , and let the automorphism  $T$  of  $G_0$  be  $T(1, t, t^2) = (1, t^2, t)$ . Our hypothesis is verified, thus giving us a 3-group  $(s_0, ts_0, t^2s_0)$  of order three. We obtain directly  $(ts_0)^{-1}s_0(ts_0) = t^2s_0$ ,  $s_0^{-1}ts_0s_0 = t^2s_0$ ,  $s_0^{-1}t^2s_0s_0 = ts_0$ , proving that none of the three elements of the 3-group are invariant under the 3-group.

This theorem may be used to determine all finite abstract polyadic groups of given small order. In this connection we have as an immediate consequence of the preceding theorem the following. *A necessary and sufficient condition that two  $m$ -groups  $G'$  and  $G''$  be simply isomorphic is that a simple isomorphism can be set up between their associated 2-groups  $G'_0$  and  $G''_0$ , and an element  $s'_0$  of  $G'$  made to correspond to an element  $s''_0$  of  $G''$ , so that  $(s'_0)^{m-1}$  in  $G'_0$  corresponds to  $(s''_0)^{m-1}$  in  $G''_0$ , and  $s'_0$  and  $s''_0$  transform  $G'_0$  and  $G''_0$  respectively so that corresponding elements go over into corresponding elements.* We postpone the application of these theorems even to our modest determination of the polyadic groups of the first three orders until our detailed study of cyclic polyadic groups of finite order gives us some basis for comparison of polyadic groups.

However, one result of some theoretical interest emerges immediately. From our general determination theorem, it follows that the number of  $m$ -adic groups with  $g$  given symbols as elements is no greater than the number of 2-groups on  $g$  other given symbols as elements times  $g$  times the largest number of automorphisms a 2-group of order  $g$  can have. We may therefore conclude that the number of abstract  $m$ -adic groups of given finite order  $g$  is a bounded function of  $m$ .



## II. FINITE POLYADIC GROUPS

A.  $m$ -ADIC SUBSTITUTIONS AND SUBSTITUTION GROUPS

9. **The symmetric  $m$ -adic substitution group of degree  $n$ .** An ordinary substitution, finite or infinite, may be considered to be a 1-1 correspondence between the members of a class  $\Gamma$  and the members of the same class. Let now  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  be an ordered sequence of  $m-1$  equivalent classes. By an  $m$ -adic substitution on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  we shall mean a transformation which in 1-1 fashion carries the members of  $\Gamma_1$  into those of  $\Gamma_2$ , of  $\Gamma_2$  into those of  $\Gamma_3, \dots$ , of  $\Gamma_{m-1}$  into those of  $\Gamma_1$ <sup>(38)</sup>. Symbolically we shall write  $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$ . Intrinsically, therefore, the  $\Gamma$ 's really enter into an  $m$ -adic substitution as a cycle, with  $\Gamma_1$  following  $\Gamma_{m-1}$ . If  $s_1$  and  $s_2$  represent two  $m$ -adic substitutions on the same sequence of  $\Gamma$ 's, we may as usual refer to  $s_1 s_2$ , the product of  $s_1$  and  $s_2$ , that is, the transformation equivalent to performing  $s_1$  followed by  $s_2$ . But in general, for  $m > 2$ , the product of two  $m$ -adic substitutions will not be an  $m$ -adic substitution on the given sequence of  $\Gamma$ 's, for it will transform  $\Gamma_1$  into  $\Gamma_3$ , instead of  $\Gamma_2$ . On the other hand, the product of  $m$   $m$ -adic substitutions on the  $m-1$   $\Gamma$ 's will again transform  $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$ , and hence we can expect to have  $m$ -adic groups of  $m$ -adic substitutions<sup>(39)</sup>. We can likewise expect to have  $m$ -adic groups of  $\mu$ -adic substitutions provided  $\mu-1$  is a divisor of  $m-1$ . However, by  $m$ -adic substitution group we shall understand the former, that is, a set of  $m$ -adic substitutions, all on the same sequence of  $\Gamma$ 's, and forming an  $m$ -adic group under the product of  $m$  substitutions as operation<sup>(40)</sup>.

When the  $\Gamma$ 's are mutually exclusive, an  $m$ -adic substitution can be given by an ordinary substitution where the one class  $\Gamma$  is the logical sum of the given  $\Gamma$ 's. On the other hand, when the  $\Gamma$ 's have common elements, an  $m$ -adic substitution cannot in general be thus considered, since one and the same element may be transformed into different elements according to the  $\Gamma_i$  of which it is considered to be a member. We shall restrict our attention to the former case<sup>(41)</sup>. But our results will be foreshadowed not by considering the resulting

<sup>(38)</sup> Our language is that of transformation; that is, we shall say " $a$  is carried into  $b$ " where the language of substitution would say " $a$  is replaced by  $b$ ."

<sup>(39)</sup> On the other hand, the product of  $m$   $m$ -adic substitutions not all on the same sequence of  $\Gamma$ 's will "usually" fail to be an  $m$ -adic substitution for any sequence of  $\Gamma$ 's. Hence the straight-laced definition following.

<sup>(40)</sup> The following generalization of ordinary substitution likewise suggests itself in connection with the *schar* concept. For but two equivalent classes  $\Gamma_1, \Gamma_2$ , consider transformations which in 1-1 fashion carry the elements of  $\Gamma_1$  into those of  $\Gamma_2$ . If  $A, B, C$  are three such transformations, then  $AB^{-1}C$  is also such a transformation. Note that here the product of two such transformations does not, in general, even exist.

<sup>(41)</sup> For simplicity. If each member  $a$  of  $\Gamma_i$  is replaced by the couple  $(i, a)$ ,  $\Gamma$ 's not mutually exclusive become mutually exclusive, and it is then readily seen when results obtained for mutually exclusive  $\Gamma$ 's hold for arbitrary  $\Gamma$ 's. Actually, our results were first obtained for arbitrary  $\Gamma$ 's. But that they are so little affected by the overlapping or nonoverlapping of the  $\Gamma$ 's indicates that we have left wholly unexplored the more interesting part of the complete theory.

$m$ -adic substitutions special types of ordinary substitutions, but generalizations of ordinary substitutions, reducing to the latter when  $m=2$ .

We further restrict our attention to the case where the  $\Gamma$ 's are finite classes, and hence consist each of the same finite number of members  $n$ . The analogy with an ordinary substitution will be furthered by saying that the  $m$ -adic substitution is then of *degree*  $n$ . Let then the members of  $\Gamma_1$  be symbolized  $a_{11}, a_{12}, \dots, a_{1n}$ , of  $\Gamma_2, a_{21}, a_{22}, \dots, a_{2n}$ ,  $\dots$  of  $\Gamma_{m-1}, a_{(m-1)1}, a_{(m-1)2}, \dots, a_{(m-1)n}$ . Corresponding to the primitive mode of writing ordinary substitutions we have the following form for any  $m$ -adic substitution on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ :

$$\begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ a_{2j'_1} & a_{2j'_2} & \dots & a_{2j'_n} \\ \cdot & \cdot & \dots & \cdot \\ a_{(m-1)j^{(m-1)}_1} & a_{(m-1)j^{(m-1)}_2} & \dots & a_{(m-1)j^{(m-1)}_n} \\ a_{1j^{(m)}_1} & a_{1j^{(m)}_2} & \dots & a_{1j^{(m)}_n} \end{array}$$

where the  $i$ th row is some permutation of  $(a_{i1}a_{i2} \dots a_{in})$  except for  $i=m$ , when it is a permutation of the first row, and each letter is carried into the one immediately below it by the substitution. If, as suggested above, we consider our  $m$ -adic substitution an ordinary substitution on all the letters  $a_{ij}$ , it can also be written in standard form as a product of cycles on different letters. In that case, each cycle will have a multiple of  $m-1$  letters, these letters cyclically running through the  $m-1$   $\Gamma$ 's.

Since an  $m$ -adic substitution of degree  $n$  is thus determined by  $m-1$  independent permutations of  $n$  elements each, we thus see that there are  $(n!)^{m-1}$   $m$ -adic substitutions of degree  $n$ , the sequence of  $\Gamma$ 's being understood given. Observe again that if  $s_1, s_2, \dots, s_m$  are  $m$ -adic substitutions on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , their products  $s_1s_2 \dots s_m$  is also an  $m$ -adic substitution on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ . In detail,  $s_1$  will carry  $a_{ij}$  into some  $a_{(i+1)j'}$ ,  $s_2$  will carry  $a_{(i+1)j'}$  into some  $a_{(i+2)j''}$ ,  $\dots$ , and  $s_m$  a resulting  $a_{ij^{(m-1)}}$  into  $a_{(i+1)j^{(m)}}$ . Hence  $s_1s_2 \dots s_m$  carries  $a_{ij}$  into  $a_{(i+1)j^{(m)}}$  as required. It then easily follows that the  $(n!)^{m-1}$   $m$ -adic substitutions of degree  $n$  constitute an  $m$ -group under the operation  $s_1s_2 \dots s_m$ . While the corresponding result holds good apart from our hypothesis of finite mutually exclusive  $\Gamma$ 's, for the present case it suffices to reinterpret our  $m$ -adic substitutions as ordinary substitutions. Condition 2 for an  $m$ -group then follows from the associative law for the multiplication of ordinary substitutions. As for condition 1, the case where all  $s$ 's but  $s_{m+1}$  in  $s_1s_2 \dots s_m = s_{m+1}$  are given  $m$ -adic substitutions has been taken care of. And if all but  $s_i$  are given  $m$ -adic substitutions,  $1 \leq i \leq m$ , by letting  $s_i$  run through the  $(n!)^{m-1}$  possible  $m$ -adic substitutions,  $s_1s_2 \dots s_m$  must do the same, and hence equals  $s_{m+1}$  for one and only one  $m$ -adic substitution  $s_i$ .

We shall call this  $m$ -group of order  $(n!)^{m-1}$  the  *$m$ -adic symmetric group of*

degree  $n$ . It clearly becomes the ordinary symmetric group of degree  $n$  when  $m=2$ . As in the case of ordinary substitution groups, every  $m$ -adic substitution group on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , or briefly of degree  $n$ , will be a subgroup of the  $m$ -adic symmetric group of degree  $n$ . It readily follows that the necessary and sufficient condition that a finite set of  $m$ -adic substitutions all on the same sequence of  $\Gamma$ 's form an  $m$ -adic substitution group is that the product of any  $m$  substitutions in the set be in the set.

Of special interest are those  $m$ -adic substitutions of degree  $n$  in which the last row is an exact repetition of the first row. There are clearly  $(n!)^{m-2}$  such substitutions. If  $s$  be such a substitution,  $s^{m-1}$  clearly carries each letter into itself, and hence  $s^m = s$ . Conversely, if  $s^m = s$ ,  $s$  must be such a substitution. According to a definition already given,  $s$  is then of  $m$ -adic order one. The unit class with  $s$  as sole member therefore itself constitutes an  $m$ -adic substitution group of order one. Hence the  $m$ -adic symmetric group of degree  $n$  has  $(n!)^{m-2}$  first order elements, and correspondingly  $(n!)^{m-2}$  subgroups of order one. For  $m=2$  these become the sole identity of the group.

10.  $2^{m-1}$ -fold classification of  $m$ -adic substitutions; the  $m$ -adic alternating groups. The classic theory of positive and negative substitutions involves the use of the determinant

$$\Delta = \begin{vmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{vmatrix}$$

which is left invariant under every positive substitution on the letters  $a_1, a_2, \dots, a_n$ , and is transformed into its negative under every negative substitution on those letters. We generalize this theory by the same means.

We now form the  $m-1$  determinants  $\Delta_1, \Delta_2, \dots, \Delta_{m-1}$ , where  $\Delta_i$  is the determinant  $\Delta$  for the  $n$  letters  $a_{i1}, a_{i2}, \dots, a_{in}$  of  $\Gamma_i$ , and transform them accordingly to a given  $m$ -adic substitution

$$\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{2j_1'} & a_{2j_2'} & \cdots & a_{2j_n'} \\ \cdot & \cdot & \cdots & \cdot \\ a_{(m-1)j_1^{(m-1)}} & a_{(m-1)j_2^{(m-1)}} & \cdots & a_{(m-1)j_n^{(m-1)}} \\ a_{1j_1^{(m)}} & a_{1j_2^{(m)}} & \cdots & a_{1j_n^{(m)}} \end{array}$$

If in the  $i$ th row each letter  $a_{ij}$  is rewritten  $a_{(i+1)j^{(42)}}$ , then the new  $i$ th row together with the old  $(i+1)$ -st row defines an ordinary substitution on the letters of the  $(i+1)$ -st row. The transform of  $\Delta_i$  under the  $m$ -adic substitu-

<sup>(42)</sup>  $a_{1j}$ , when  $i=m-1$ . Likewise, below,  $\Delta_{i+1}$  is  $\Delta_1$  when  $i=m-1$ .

tion is clearly the transform of  $\Delta_{i+1}$  under this ordinary substitution, and hence is  $\Delta_{i+1}$ , or its negative, according as this ordinary substitution is positive or negative. We therefore have under the  $m$ -adic substitution

$$\Delta_1 \rightarrow \delta_1 \Delta_2, \Delta_2 \rightarrow \delta_2 \Delta_3, \dots, \Delta_{m-1} \rightarrow \delta_{m-1} \Delta_1, \quad \delta_1, \delta_2, \dots, \delta_{m-1} = \pm 1.$$

With each  $m$ -adic substitution there is thus associated a sequence of  $m-1$  numbers  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$  whose values are  $+1$  or  $-1$ . Clearly, when  $n > 1$ , an  $m$ -adic substitution of degree  $n$  can be written down for every possible assignment of values to the  $\delta$ 's. The  $m$ -adic substitutions of degree  $n$ ,  $n > 1$ , thus fall into  $2^{m-1}$  mutually exclusive classes corresponding to the  $2^{m-1}$  possible  $\delta$ -sequences  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ .

Given  $m$   $m$ -adic substitutions  $s_i$ , with the corresponding  $\delta$ -sequences  $[\delta_{i1}, \delta_{i2}, \dots, \delta_{i(m-1)}]$ , the  $m$ -adic substitution  $s_1 s_2 \dots s_m$  has a  $\delta$ -sequence  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$  which depends only on the  $\delta$ -sequences of the  $s_i$ 's. In fact, by following through the effect of the succession of substitutions  $s_1, s_2, \dots, s_m$  on the determinants  $\Delta_1, \Delta_2, \dots, \Delta_{m-1}$ , we obtain the following equations for determining  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ :

$$\delta_1 = \delta_{11} \delta_{22} \dots \delta_{(m-1)(m-1)} \delta_{m1},$$

$$\delta_2 = \delta_{12} \delta_{23} \dots \delta_{(m-1)1} \delta_{m2},$$

$$\dots \dots \dots$$

$$\delta_{m-1} = \delta_{1(m-1)} \delta_{21} \dots \delta_{(m-1)(m-2)} \delta_{m(m-1)}.$$

Now let  $K$  be the class of the  $2^{m-1}$  possible  $\delta$ -sequences. If  $\sigma_1, \sigma_2, \dots, \sigma_m$  are any  $m$  such  $\delta$ -sequences, and  $\sigma$  is the  $\delta$ -sequence obtained from  $\sigma_1, \sigma_2, \dots, \sigma_m$  in accordance with the above equations, an  $m$ -adic operation  $k(\sigma_1 \sigma_2 \dots \sigma_m)$  is determined such that  $\sigma = k(\sigma_1 \sigma_2 \dots \sigma_m)$ . It is then readily shown that  $K$  constitutes an  $m$ -group under  $k$ . In fact, condition 1 for an  $m$ -group is immediately verified by referring to the above equations. And condition 2 follows from the associative law for  $m$ -adic substitutions, and the fact that if  $s_1, s_2, \dots, s_m$  are substitutions corresponding to  $\sigma_1, \sigma_2, \dots, \sigma_m$  respectively,  $s_1 s_2 \dots s_m$  corresponds to  $k(\sigma_1 \sigma_2 \dots \sigma_m)$ . We shall call this  $m$ -group of order  $2^{m-1}$  the *complete  $m$ -adic  $\delta$ -group*<sup>(43)</sup>.

Consider now any  $m$ -adic substitution group of degree  $n$  and form the class  $K'$  of  $\delta$ -sequences corresponding to its members. Since the product of any  $m$  substitutions of the group is in the group, the  $k$  product of any  $m$   $\delta$ -sequences in  $K'$  will be in  $K'$ . As  $K'$  is a subclass of the class of members of the complete  $m$ -adic  $\delta$ -group, and the latter is finite, this suffices to prove

<sup>(43)</sup> Actually, then, we have established a homomorphism between the symmetric  $m$ -adic substitution group of degree  $n$ ,  $n > 1$ , and this complete  $m$ -adic  $\delta$ -group. The rest of this section could then largely have been given as a consequence of our general results on homomorphisms between  $m$ -adic groups, as could indeed the very fact that  $K$  is an  $m$ -group under  $k$ . In the generalization of this section occurring in the last section of our paper full use will be made of the concept of homomorphism.

the following. The  $\delta$ -sequences corresponding to the members of any  $m$ -adic substitution group of degree  $n$  form the complete  $m$ -adic  $\delta$ -group, or a subgroup thereof.

By means of the above equations for the  $k$  operation we readily prove, as for ordinary substitutions, that every  $m$ -adic substitution group of degree  $n$  has the same number of substitutions for each  $\delta$ -sequence in the corresponding " $\delta$ -subgroup"<sup>(44)</sup>. In fact, let  $s_m$  and  $s_{m+1}$  be any two substitutions in the group corresponding to any two given  $\delta$ -sequences  $\sigma_m$  and  $\sigma_{m+1}$  of the corresponding  $\delta$ -subgroup, and choose  $s_1, \dots, s_{m-1}$  so that  $s_1 \dots s_{m-1} s_m = s_{m+1}$ . If now we let  $s_m$  run through all the substitutions in the group corresponding to  $\sigma_m$ ,  $s_{m+1}$  assumes an equal number of values in the group all corresponding to  $\sigma_{m+1}$ . Hence there are at least as many substitutions in the group corresponding to one  $\delta$ -sequence as to another, and consequently, by reciprocal reasoning, the same number. Since the order of the complete  $m$ -adic  $\delta$ -group is  $2^{m-1}$ , that of a subgroup thereof must be of the form  $2^u$ <sup>(45)</sup>. From the above result it follows that the order of an  $m$ -adic substitution group is a multiple of the order of its  $\delta$ -subgroup. We therefore have as a corollary of the above result every  $m$ -adic substitution group of odd order has a  $\delta$ -subgroup of order one, that is, all of its substitutions correspond to one and the same  $\delta$ -sequence.

Applied to the symmetric group itself, the above result shows that the  $2^{m-1}$  mutually exclusive classes into which the  $m$ -adic symmetric group of degree  $n$  is divided all have the same number of members. Now given any subgroup of the complete  $m$ -adic  $\delta$ -group, form the class  $C'$  of all the  $m$ -adic substitutions of degree  $n$  corresponding to each  $\delta$ -sequence in the given  $\delta$ -subgroup. The product of any  $m$  substitutions in  $C'$  will therefore be in  $C'$ . Hence the members of  $C'$  form a subgroup of the symmetric group. By analogy with ordinary groups we shall call it an  $m$ -adic alternating group. Consequently, there are as many  $m$ -adic alternating groups of degree  $n$ ,  $n > 1$ , as there are subgroups of the complete  $m$ -adic  $\delta$ -group, each alternating group consisting of all the substitutions of the symmetric group with  $\delta$ -sequences in the corresponding  $\delta$ -subgroup. We may now further state that there is a one-many correspondence between the  $m$ -adic  $\delta$ -subgroups and  $m$ -adic substitution groups of degree  $n$ ,  $n > 1$ , that is, between the class consisting of the complete  $m$ -adic  $\delta$ -group and its subgroups, and the  $m$ -adic symmetric group of degree  $n$  and its subgroups; and this correspondence is preserved under the relation "group or subgroup of."

For  $m = 2$  the complete  $\delta$ -group is the cyclic group of order 2, and its sole subgroup, the identity, corresponds to the sole ordinary alternating group of degree  $n$ <sup>(46)</sup>. For  $m = 3$  the complete  $\delta$ -group is of order 4. By direct calcula-

<sup>(44)</sup> We shall use the phrase  $\delta$ -subgroup to cover the complete  $\delta$ -group as well.

<sup>(45)</sup> By Lagrange's theorem for polyadic groups—proved in §4.

<sup>(46)</sup> van der Waerden has already noted the homomorphism between any substitution group having at least one odd substitution and this cyclic group of order two.



tion we find it to possess exactly four subgroups, that is, with classes of elements  $([+1, +1])$ ,  $([-1, -1])$ ,  $([+1, +1], [-1, -1])$ ,  $([+1, -1], [-1, +1])$ . Hence, there are exactly four triadic alternating groups of degree  $n$ ,  $n > 1$ .

Thanks to B. P. Gill, we are able to determine the  $m$ -adic alternating groups of degree  $n$  for arbitrary  $m$ . For this purpose it is essential to obtain a suitable representation of the associated ordinary group of the complete  $m$ -adic  $\delta$ -group. The ideas leading up to this are of more general application, and hence at least part of the following digression.

**11. Associated and containing ordinary groups; commutative  $m$ -adic substitutions.** The substitutions of an  $m$ -adic substitution group  $G$  of degree  $n$ , considered as ordinary substitutions on  $(m-1)n$  letters, generate an ordinary substitution group which satisfies our definition of a containing group of  $G$ . With  $G$  thus an  $m$ -adic group of  $m$ -adic substitutions, this containing group will be of index  $m-1$ , and hence simply isomorphic with the abstract containing group  $G^*$  of  $G$ . We shall therefore use it throughout to represent  $G^*$ , and for simplicity symbolize it  $G^*$ . We may likewise refer to the associated group of  $G$  with respect to this containing group as  $G_0$ .

In the terminology of §6, the  $i$ th coset of  $G^*$  consists of the products of  $i$  elements of  $G$ . To avoid duplication, it will be convenient henceforth to assume that  $1 \leq i \leq m-1$ . Since each substitution in  $G$  transforms  $\Gamma_1 \rightarrow \Gamma_2$ ,  $\Gamma_2 \rightarrow \Gamma_3$ ,  $\dots$ ,  $\Gamma_{m-1} \rightarrow \Gamma_1$ , it follows that the  $i$ th coset of  $G^*$  consists of transformations which in 1-1 fashion carry the members of each  $\Gamma_i$  into those of  $\Gamma_{i+i}$ ,  $j+i$  reduced modulo  $m-1$  if need be. We may therefore call these substitutions of  $G^*$  the  $i$ -ads of  $G$ . In particular,  $G_0$ , which consists of the  $(m-1)$ -ads of  $G$  in  $G^*$ , consists of transformations which transform each  $\Gamma_i$  into itself. Each  $(m-1)$ -ad of  $G$  thus appears in  $G^*$  as the product of  $m-1$  ordinary substitutions, each of these ordinary substitutions being on the letters of a single  $\Gamma$ . We have incidentally verified that  $G^*$  is of index  $m-1$ .

Considered as ordinary substitution groups on  $(m-1)n$  letters we see that for  $m > 2$ ,  $G^*$  is imprimitive with systems of imprimitivity  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , while  $G_0$  is intransitive with the letters in each  $\Gamma$  carried into letters of the same  $\Gamma$  only, by every substitution of  $G$ . If then for each  $\Gamma$  we separate from each substitution in  $G_0$  the substitution involving only the letters of that  $\Gamma$ , there results an ordinary substitution group on the letters of that  $\Gamma$ . We shall symbolize these  $m-1$  groups on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  by  $G'_0, G''_0, \dots, G_0^{(m-1)}$  respectively, and call them the *associated constituent groups* of the  $m$ -adic substitution group  $G$ . It is then significant that the *associated constituent groups of an  $m$ -adic substitution group are conjugate ordinary groups*. In fact, recall that  $G_0$  is an invariant subgroup of  $G^*$ , and hence is invariant under every  $m$ -adic substitution  $s$  in  $G$ . Now  $s$  carries the letters of each  $\Gamma_i$  into those of  $\Gamma_{i+1}$ . Hence, when the substitutions of  $G_0$  are transformed by  $s$ , the components of these substitutions on the letters of  $\Gamma_i$  be-



come the components of the same class of substitutions on the letters of  $\Gamma_{i+1}$ . We thus have specifically

$$s^{-1}G'_0s = G''_0, \quad s^{-1}G''_0s = G'''_0, \quad \dots, \quad s^{-1}G_0^{(m-1)}s = G'_0,$$

for every  $s$  in  $G$ .

If  $s_1$  and  $s_2$  are  $m$ -adic substitutions on the same sequence of  $\Gamma$ 's, we may consider them as elements of the corresponding  $m$ -adic symmetric group. The transform of  $s_2$  under  $s_1$  is then  $s_1^{-1}s_2s_1$  in the notation of the containing group of the symmetric group, and hence may be obtained by the ordinary rule for transforming substitutions. Restated for our primitive mode of representing  $m$ -adic substitutions, this rule becomes the following. Replace each letter in  $s_2$  by the letter immediately under it in  $s_1$  and rewrite in standard form. Thus, to illustrate, let

$$\begin{array}{ccccc} a_{11}a_{12}a_{13} & & a_{11}a_{12}a_{13} & & a_{22}a_{23}a_{21} & a_{11}a_{12}a_{13} \\ s_2 = a_{22}a_{21}a_{23}, & s_1 = a_{22}a_{23}a_{21}; & s_1^{-1}s_2s_1 = a_{13}a_{12}a_{11} = & a_{23}a_{21}a_{22}. \\ a_{11}a_{13}a_{12} & & a_{13}a_{11}a_{12} & & a_{22}a_{21}a_{23} & a_{12}a_{11}a_{13} \end{array}$$

Actually, the result before it is rewritten defines the transform equally well; for, as stated before, it is really the cycle, rather than the sequence, of  $\Gamma$ 's that is significant.

If  $s_2$  is invariant under  $s_1$ , then  $s_1$  and  $s_2$  are commutative; and conversely. The problem of determining all  $m$ -adic substitutions  $s$ , commutative with a given  $m$ -adic substitution  $s_1$  of degree  $n$ , and on the same  $\Gamma$ 's, is best treated by writing the substitutions in ordinary cycle form. We recall that the number of letters in each cycle is then a multiple of  $m-1$ . If  $s_1$  consists of a single cycle, the ordinary substitutions  $r$ , on the  $(m-1)n$  letters of  $s_1$ , which are commutative with  $s_1$ , are the  $(m-1)n$  ordinary powers of  $s_1$ . Of these exactly  $n$ , i.e., those of the form  $s_1^{k(m-1)+1}$ , are  $m$ -adic substitutions on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ . We shall later call these the  $m$ -adic powers of  $s_1$ , i.e., the elements of the  $m$ -adic group generated by  $s_1$ . Hence, the only  $m$ -adic substitutions on the  $\Gamma$ 's of  $s_1$ , commutative with the single cycle  $m$ -adic substitution  $s_1$  of degree  $n$ , are the  $n$   $m$ -adic powers of  $s_1$ . We now have no difficulty in paraphrasing the corresponding argument for ordinary substitutions, and obtain the following results. If  $s_1$  consists of  $\lambda$  cycles with numbers of letters  $(m-1)n_1, (m-1)n_2, \dots, (m-1)n_\lambda$  no two of which are equal, the  $m$ -adic substitutions  $s$ , on the  $\Gamma$ 's of  $s_1$ , commutative with  $s_1$ , are the  $n_1n_2 \dots n_\lambda$  products of the  $m$ -adic powers of the several cycles. And, if  $s_1$  consists of  $k$  equal cycles of  $(m-1)\nu$  letters each, the  $m$ -adic substitutions  $s$  on the  $\Gamma$ 's of  $s_1$  commutative with  $s_1$  are  $\nu^k k!$  in number, there being  $\nu^k$  such  $m$ -adic substitutions for each of the  $k!$  possible permutations of the  $k$  cycles. Clearly in any case, the  $m$ -adic substitutions  $s$ , commutative with an  $m$ -adic substitution  $s_1$ , and on the  $\Gamma$ 's of  $s_1$ , constitute an  $m$ -adic substitution group.

**12. Further study of the complete  $m$ -adic  $\delta$ -group and  $m$ -adic alternating groups.** The ideas of the preceding section enable us to clear up a certain difficulty in our presentation of the  $2^{m-1}$ -fold classification of  $m$ -adic substitutions and in its consequences. Observe that whereas the  $\Gamma_i$ 's are mere classes, the determinants  $\Delta_i$  assume the letters in each  $\Gamma_i$  arranged in a sequence. The  $\delta$ -sequence associated with a given  $m$ -adic substitution  $s$  will therefore in general depend not only on  $s$  but also on the original ordering of the letters in the  $\Gamma_i$ 's. However, we shall see that the same  $2^{m-1}$  classes are obtained no matter what ordering is assumed, only their description by  $\delta$ -sequences being thus affected.

Actually, this ordering is equivalent to a first order  $m$ -adic substitution  $s_0$  which carries each  $a_{ij}$  into  $a_{(i+1)j}$ ,  $i+1$  being replaced by 1 when  $i=m-1$ . Let us then write the  $m$ -adic symmetric group in coset form with arbitrary element  $s = ts_0$ .  $t$  is then in the form  $t^{(1)}t^{(2)} \cdots t^{(m-1)}$  where  $t^{(i)}$  is an ordinary substitution on the letters of  $\Gamma_i$ . If now we associate with  $t$  the  $\epsilon$ -sequence  $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$ , where  $\epsilon_i$  is  $+1$  or  $-1$  according as  $t^{(i)}$  is a positive or negative substitution, we see from the effect on the determinants  $\Delta_i$  that the  $\epsilon$ -sequence of  $t$  is identical with the  $\delta$ -sequence of  $s$ . Let then  $s_1 = t_1s_0$  and  $s_2 = t_2s_0$  have the same  $\delta$ -sequence, and hence  $t_1$  and  $t_2$  the same  $\epsilon$ -sequence. Then  $s_1s_2^{-1} = t_1t_2^{-1}$  will have an  $\epsilon$ -sequence  $(+1, +1, \dots, +1)$ , i.e., will be the product of positive substitutions only. Conversely, if  $s_1s_2^{-1}$  is the product of positive substitutions only, the corresponding  $t_1$  and  $t_2$  must have the same  $\epsilon$ -sequence, and  $s_1$  and  $s_2$  the same  $\delta$ -sequence. Hence,  $s_1$  and  $s_2$  belong to the same one of the  $2^{m-1}$  classes of  $m$ -adic substitutions when and only when  $s_1s_2^{-1}$  is the product of positive substitutions on the letters of the several  $\Gamma$ 's. As this criterion is independent of  $s_0$ , the intrinsic character of our classification has been demonstrated.

The  $\epsilon$ -sequences may be used to obtain a concrete representation of the associated ordinary group of the complete  $m$ -adic  $\delta$ -group. More generally, consider the containing group of the  $m$ -adic symmetric group of degree  $n$ ,  $n > 1$ . Since each  $i$ -ad  $R$  thereof is the product of  $i$   $m$ -adic substitutions,  $R$  will transform the  $\Delta$ 's according to some scheme

$$\Delta_1 \rightarrow \eta_1 \Delta_{i+1}, \Delta_2 \rightarrow \eta_2 \Delta_{i+2}, \dots; \Delta_{m-1} \rightarrow \eta_{m-1} \Delta_i, \quad \eta_1, \eta_2, \dots, \eta_{m-1} = \pm 1.$$

With  $R$  we may thus associate the  $\eta$ -sequence (with subscript)  $\{\eta_1, \eta_2, \dots, \eta_{m-1}\}_i$ . If  $R_1$  thus corresponds to  $\{\eta'_1, \eta'_2, \dots, \eta'_{m-1}\}_i$ ,  $R_2$  to  $\{\eta''_1, \eta''_2, \dots, \eta''_{m-1}\}_i$ ,  $R_1R_2$  will correspond to  $\{\eta'_1\eta''_{i+1}, \eta'_2\eta''_{i+2}, \dots, \eta'_{m-1}\eta''_{i+m}\}_{i+i}$ , subscripts being reduced modulo  $m-1$  if need be. It follows that the containing group of the  $m$ -adic symmetric group is homomorphic to the resulting complete  $\eta$ -group (with subscript). Now with  $i=1$ , the  $\eta$ -sequence is nothing more than the  $\delta$ -sequence of the corresponding  $m$ -adic substitution. From the way in which our operations were obtained it follows that the complete  $\eta$ -group may be considered a containing group, of index  $m-1$ ,

indeed, of the complete  $m$ -adic  $\delta$ -group. The associated group of the complete  $m$ -adic  $\delta$ -group will then be composed of the  $\eta$ -sequences whose subscript is  $m-1$ . But going back to the  $\Delta$ 's we see that these  $\eta$ -sequences are then actually the  $\epsilon$ -sequences of the corresponding  $(m-1)$ -ads of  $m$ -adic substitutions. Under this representation, therefore, the operation of the associated group of the complete  $m$ -adic  $\delta$ -group, i.e., of the complete  $\epsilon$ -group as we shall call it, becomes

$$(\epsilon'_1, \epsilon'_2, \dots, \epsilon'_{m-1})(\epsilon''_1, \epsilon''_2, \dots, \epsilon''_{m-1}) = (\epsilon'_1\epsilon''_1, \epsilon'_2\epsilon''_2, \dots, \epsilon'_{m-1}\epsilon''_{m-1})^{(47)}.$$

We therefore see that the complete  $\epsilon$ -group is an ordinary abelian group of order  $2^{m-1}$ . Since each  $\epsilon$  is  $\pm 1$ , its elements other than the identity are all of order two, so that it is indeed of type  $(1, 1, \dots, 1)$ .

The complete  $m$ -adic  $\delta$ -group is therefore semi-abelian. As it is readily seen to be non-abelian whenever  $m > 2$ , it follows that it then has no invariant element. More specifically, the transform of  $[\delta_1, \delta_2, \delta_3, \dots, \delta_{m-1}]$  by  $[\delta'_1, \delta'_2, \delta'_3, \dots, \delta'_{m-1}]$  is easily found, via the complete  $\eta$ -group, to be

$$[\delta'_{m-1}\delta_{m-1}\delta'_1, \delta'_1\delta_1\delta'_2, \delta'_2\delta_2\delta'_3, \dots, \delta'_{m-2}\delta_{m-2}\delta'_{m-1}].$$

The condition for invariance is then easily rewritten  $\delta_1\delta'_1 = \delta_2\delta'_2 = \delta_3\delta'_3 = \dots = \delta_{m-1}\delta'_{m-1}$ . It follows that there are exactly two  $\delta$ -sequences leaving any given  $\delta$ -sequence  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$  invariant, namely,  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$  and  $[-\delta_1, -\delta_2, \dots, -\delta_{m-1}]$ .

The present and succeeding paragraph presuppose a partial reading of the later §21 and §22. We have observed that except for the identity the elements of the complete  $\epsilon$ -group are all of order two. While it follows therefrom that the elements of the complete  $m$ -adic  $\delta$ -group are of no other  $m$ -adic orders than one or two, we find directly that exactly half of them are of order one, half of order two. Thus, if  $\sigma$  is the  $\delta$ -sequence  $[\delta_1, \delta_2, \dots, \delta_{m-1}]$ , and  $\delta_0 = \delta_1\delta_2 \dots \delta_{m-1}$ , then, with  $k$  as in §10, we find  $k(\sigma\sigma \dots \sigma) = [\delta_0\delta_1, \delta_0\delta_2, \dots, \delta_0\delta_{m-1}]$ ,  $k(\sigma\sigma \dots k(\sigma\sigma \dots \sigma)) = [\delta_1, \delta_2, \dots, \delta_{m-1}]$ . Hence, the  $m$ -adic order of a  $\delta$ -sequence is one or two according as the product of its  $\delta$ 's is  $+1$  or  $-1$ .

The cyclic subgroups of the complete  $m$ -adic  $\delta$ -group are therefore of orders one or two, there being  $2^{m-2}$  first order subgroups, and, for  $m > 2$ ,  $2^{m-2}$  or  $2^{m-3}$  cyclic second order subgroups according as  $m$  is even or odd. Our result on the  $\delta$ -sequences leaving a given  $\delta$ -sequence invariant, coupled with the easily verified fact that an  $m$ -group of order two must be abelian, leads to the result that the complete  $m$ -adic  $\delta$ -group has exactly  $2^{m-2}$  second

<sup>(47)</sup> Actually, by a slight change in point of view, the transformation of the  $\Delta$ 's resulting from an  $m$ -adic substitution can be considered an  $m$ -adic linear transformation in one variable in the sense of our later §35. The present and several other formulas, derived independently in the present section, would then become special cases of the formulas of §35.

order subgroups for  $m > 2$ . Hence, when  $m$  is odd, half of them are non-cyclic<sup>(48)</sup>.

We turn now to the determination of all the subgroups of the complete  $m$ -adic  $\delta$ -group, and consequently, the determination of all  $m$ -adic alternating groups. Since the complete  $m$ -adic  $\delta$ -group is semi-abelian, all of its elements transform a given  $\epsilon$ -sequence  $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$  into the same  $\epsilon$ -sequence. As before, we can employ the operation of the complete  $\eta$ -group, and thus find the unique transform of  $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$  under every  $\delta$ -sequence to be  $(\epsilon_{m-1}, \epsilon_1, \epsilon_2, \dots, \epsilon_{m-2})$ . Now if  $H$  is a subgroup of the complete  $m$ -adic  $\delta$ -group, its associated ordinary group  $H_0$  must be a subgroup of the complete  $\epsilon$ -group invariant under  $H$ . Hence  $H_0$  can only be such a subgroup of the complete  $\epsilon$ -group that if  $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-2}, \epsilon_{m-1})$  is in the subgroup,  $(\epsilon_{m-1}, \epsilon_1, \epsilon_2, \dots, \epsilon_{m-2})$  also is in the subgroup. The determination of these "admissible" subgroups of the complete  $\epsilon$ -group is the only difficult part of our problem. It was carried through independently by Gill; but he later found that his solution followed essentially the lines of the general theory of the "Verallgemeinerte Abelsche Gruppen," abbreviated V.A.G., as given by Otto Haupt in the second volume of his *Algebra*<sup>(49)</sup>.

Following Gill we replace the two values  $+1, -1$  by  $0, 1$  respectively. If an  $\epsilon$ -sequence be thus rewritten, the dyadic operation of our complete  $\epsilon$ -group is best written in additive form, and we have

$$(\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{1(m-1)}) + (\epsilon_{21}, \epsilon_{22}, \dots, \epsilon_{2(m-1)}) \\ = (\epsilon_{11} + \epsilon_{21}, \epsilon_{12} + \epsilon_{22}, \dots, \epsilon_{1(m-1)} + \epsilon_{2(m-1)}),$$

where addition within the parentheses is modulo 2. Now let  $\phi(x)$  be any polynomial in  $x$  with coefficients 0 or 1. With  $a$  any  $\epsilon$ -sequence, a unique  $\epsilon$ -sequence  $\phi(x) \cdot a$  is determined as follows. If  $a$  is the  $\epsilon$ -sequence  $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-2}, \epsilon_{m-1})$ , let  $x \cdot a$  be the  $\epsilon$ -sequence  $(\epsilon_{m-1}, \epsilon_1, \epsilon_2, \dots, \epsilon_{m-2})$ . With  $1 \cdot a = a$ , and  $x^n \cdot a$  defined inductively through  $x^n \cdot a = x \cdot (x^{n-1} \cdot a)$ , we can define  $\phi(x) \cdot a$  as the sum of the  $\epsilon$ -sequences obtained by operating on  $a$  by the several terms of  $\phi(x)$ . We now observe two things. First, every  $\epsilon$ -sequence can be written  $\phi(x) \cdot (1, 0, \dots, 0)$ . In fact, to obtain  $(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$ , we need merely let  $\phi(x) = \epsilon_1 + \epsilon_2 x + \dots + \epsilon_{m-1} x^{m-2}$ . Secondly, with  $(0, 0, \dots, 0)$  abbreviated 0, we see that  $(1, 0, \dots, 0)$  satisfies the equation  $(x^{m-1} + 1) \cdot (1, 0, \dots, 0) = 0$ , but fails to satisfy any equation  $\phi(x) \cdot (1, 0, \dots, 0) = 0$  with  $\phi(x)$  of degree less than  $m-1$ , and not identically zero. For we have directly that  $x^{m-1} \cdot (1, 0, \dots, 0) = (1, 0, \dots, 0)$ ; while with  $\phi(x)$  of degree less than  $m-1$  our previous expression for  $\phi(x) \cdot (1, 0, \dots, 0)$  applies. Note finally that 0 and 1 constitute a field  $K$  under addition modulo 2, and multiplication. The entire theory of V.A.G.'s in general, and Theorem 3 of Haupt

<sup>(48)</sup> See §23 for the consequent structure of these second order subgroups.

<sup>(49)</sup> Otto Haupt, *Einführung in die Algebra*, Leipzig, 1929, vol. 2, pp. 617-621. The result we need is the Theorem 3 of page 620.

in particular, can then be shown to be applicable, and yield the following result.

*The admissible subgroups of the complete  $m$ -adic  $\epsilon$ -group are in 1-1 correspondence with the polynomial divisors, other than unity, of  $x^{m-1}+1$  relative to the field of coefficients  $K$ . If  $\tau(x)$  be such a divisor, and  $a = \tau(x) \cdot (1, 0, \dots, 0)$ , then the corresponding subgroup consists of all distinct  $\epsilon$ -sequences  $\phi(x) \cdot a$ .*

Actually, if  $\mu$  is the degree of  $(x^{m-1}+1)/\tau(x)$ , then  $\phi(x)$  can be restricted to degrees less than  $\mu$ , different  $\phi(x)$ 's then also giving different  $\epsilon$ -sequences. It follows that the order<sup>(60)</sup> of the corresponding subgroup is  $2^\mu$ . The subgroup corresponding to  $\tau(x)$  can also be described as consisting of all  $\epsilon$ -sequences  $b$  such that  $(x^{m-1}+1)/\tau(x) \cdot b = 0$ . It follows that these subgroups satisfy the same properties with respect to the relation of inclusion as do the subgroups of an ordinary cyclic group,  $(x^{m-1}+1)/\tau(x)$  taking the place of the order of the subgroup. Note that the unique factorization theorem applies to polynomials with coefficients in a given field. If then  $x^{m-1}+1$  is thus completely factored, the distinct divisors  $\tau(x)$  can immediately be written down. Since  $x^{m-1}+1 = (x+1)(x^{m-2} + \dots + x + 1)$  relative to  $K$ ,  $x+1$  is always one of the prime divisors of  $x^{m-1}+1$ . It can readily be shown that it is the only distinct prime divisor of  $x^{m-1}+1$ , that is, that  $x^{m-1}+1 = (x+1)^{m-1}$  relative to  $K$ , when and only when  $m-1$  is itself a power of 2. The different  $\tau(x)$ 's are then  $(x+1)$ ,  $(x+1)^2$ ,  $\dots$ ,  $(x+1)^{m-1}$ , and each corresponding subgroup contains the next.

Having determined the admissible subgroups of the complete  $\epsilon$ -group in accordance with the above theorem, it is a simple matter to find the subgroups of the complete  $\delta$ -group. We return here to our original notation. Each  $\delta$ -subgroup  $H$ , if written in coset form, will be given by  $H = H_0\sigma$ , with  $H_0$  an admissible  $\epsilon$ -subgroup,  $\sigma$  a  $\delta$ -sequence. Hence, if  $H_0\sigma$  is known to be a  $\delta$ -subgroup, its elements can immediately be found from  $H_0$  and  $\sigma$  by the relation

$$(\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1}) [\delta_1, \delta_2, \dots, \delta_{m-1}] = [\epsilon_1\delta_1, \epsilon_2\delta_2, \dots, \epsilon_{m-1}\delta_{m-1}],$$

a mere specialization of the dyadic operation of the complete  $\eta$ -group.

Since every admissible  $\epsilon$ -subgroup  $H_0$  is invariant under every  $\delta$ -sequence  $\sigma$ , it follows from an early theorem of §4 that  $H_0\sigma$  will be a  $\delta$ -subgroup for every first order  $\sigma$ , and for those second order  $\sigma$ 's for which  $\sigma^{m-1}$  is in  $H_0$ . The distinct  $\delta$ -subgroups thus arising will then be all the  $\delta$ -subgroups for a given  $H_0$ . Now when  $m$  is even, every  $\delta$ -subgroup must have at least one first order element. Hence in this case, the  $\delta$ -subgroups corresponding to  $H_0$  will be all the distinct  $H_0\sigma$ 's with  $\sigma$  a first order element. Now it is readily proved that if  $\tau = (\epsilon_1, \epsilon_2, \dots, \epsilon_{m-1})$ , the order of  $\tau\sigma$  is the same as that of  $\sigma$ , or opposite, according as  $\epsilon_0 = \epsilon_1\epsilon_2 \dots \epsilon_{m-1}$  is  $+1$  or  $-1$ , and furthermore, that the elements of  $H_0$  either all have  $\epsilon_0$  equal to  $+1$ , or exactly half have  $\epsilon_0 = +1$ ,

<sup>(60)</sup> In the ordinary sense, not that of V.A.G.'s.



half  $-1$ . Hence, if  $H_0$  is of order  $2^\mu$ ,  $H_0\sigma$ , with  $\sigma$  of first order, has  $2^\mu$  or  $2^{\mu-1}$  first order elements according as the elements of  $H_0$  have or have not  $\epsilon_0$ 's all  $+1$ . Since the distinct  $H_0\sigma$ 's with given  $H_0$  are mutually exclusive, while each of the  $2^{m-2}$  first order elements of the complete  $m$ -adic  $\delta$ -group is in some  $H_0\sigma$ , it follows that when  $m$  is even, for each admissible  $\epsilon$ -subgroup  $H_0$  of order  $2^\mu$  there are exactly  $2^{m-\mu-2}$  or  $2^{m-\mu-1}$  corresponding  $\delta$ -subgroups according as the  $\epsilon_0$ 's of the elements of  $H_0$  are, or are not, all  $+1$ .

For  $m$  odd, and given  $H_0$ , we also have these subgroups. But now there may be additional subgroups  $H_0\sigma$  with all elements of order two. Now if  $\sigma$  is of second order, the  $\epsilon$ -sequence of  $\sigma^{m-1}$  is readily seen to be  $(-1, -1, \dots, -1)$ . It follows that these additional subgroups can arise only when  $H_0$  has the element  $(-1, -1, \dots, -1)$ , while the  $\epsilon_0$ 's of all its elements are  $+1$ . But then each of the  $2^{m-2}$  second order  $\delta$ -sequences will be in one of these additional subgroups. For such an  $H_0$ , therefore, in addition to the now  $2^{m-\mu-2}$   $\delta$ -subgroups consisting wholly of first order elements, there will be  $2^{m-\mu-2}$  additional  $\delta$ -subgroups each, indeed, consisting wholly of second order elements.

Actually, the number of  $\delta$ -subgroups with given associated ordinary group  $H_0$  can be determined without explicitly writing out the elements of  $H_0$ , but merely by an inspection of the corresponding  $\tau(x)$ . Thus, we have already seen that the order of  $H_0$  is  $2^\mu$ , where  $\mu$  is the degree of  $(x^{m-1}+1)/\tau(x)$ . By means of the second description given for the subgroup  $H_0$ , it can further be shown that  $(-1, -1, \dots, -1)$  is in  $H_0$  when and only when  $(x^{m-1}+1)/\tau(x)$  has  $x+1$  for divisor; while from the first description it can be shown that the  $\epsilon_0$ 's of  $H_0$  are all  $+1$  when and only when  $\tau(x)$  has  $x+1$  for divisor. This covers all we need to know about  $H_0$ .

In particular, for  $m > 3$ , we always have the three distinct divisors of  $x^{m-1}+1$  equal to  $x^{m-1}+1$ ,  $x^{m-2}+\dots+x+1$ ,  $x+1$ . In the first case  $H_0$  is of order one, and consists of but  $(+1, +1, \dots, +1)$ , the identity. The corresponding  $\delta$ -subgroups are the first order  $\delta$ -subgroups listed above. In the second case  $H_0$  is of order two, and consists of  $(+1, +1, \dots, +1)$  and  $(-1, -1, \dots, -1)$ . It is obviously the only admissible second order  $\epsilon$ -subgroup, and hence the corresponding  $\delta$ -subgroups are all of the second order  $\delta$ -subgroups as first listed. The third subgroup, of order  $2^{m-2}$ , is again the only admissible  $\epsilon$ -subgroup of that order, and consists of all  $\epsilon$ -sequences with  $\epsilon_0$  equal to  $+1$ . Our general solution then shows that as a result there is but one  $\delta$ -subgroup of order  $2^{m-2}$  for  $m$  even, two for  $m$  odd.

Actually, the equations of §10 for the  $m$ -adic operation on  $\delta$ -sequences directly show that we always have the subgroup of order  $2^{m-2}$  consisting of all  $\delta$ -sequences with  $\delta_0 = +1$ , and for  $m$  odd also the subgroup of order  $2^{m-2}$  consisting of all  $\delta$ -sequences with  $\delta_0 = -1$ . Since the complete  $m$ -adic  $\delta$ -group is semi-abelian, all of its subgroups are semi-invariant. It is then of interest to note that the above one, or two, subgroups of order  $2^{m-2}$  are its only invariant subgroups. In fact, our formula for the transform of one  $\delta$ -sequence by an-



other shows that  $\delta_0$  is always thus left invariant; and it also shows that a  $\delta$ -sequence can always be found which transforms a given  $\delta$ -sequence into any other with the same  $\delta_0$ .

These results are immediately applicable to the corresponding alternating groups, assuming  $n > 1$ . There are thus always  $2^{m-2}$  alternating groups with substitutions forming one of the  $2^{m-1}$  classes of §10 and, for  $m > 2$ ,  $2^{m-2}$  alternating groups with substitutions forming two such classes. Passing by the general solution, we note that the conditions  $\delta_0 = +1$ , and  $\delta_0 = -1$ , correspond to an  $m$ -adic substitution considered as an ordinary substitution being positive, or negative. Hence, the only alternating groups invariant under the symmetric group are the alternating group of all positive substitutions, and, for  $m$  odd, also the alternating group of all negative substitutions. On the other hand, every alternating group is a semi-invariant subgroup of the symmetric group.

The last observation restricts the possible simplicity of  $m$ -adic alternating groups. Regarding the nonexistence of a quotient group of lower order than itself as the distinguishing mark of an ordinary simple group, we are led to define a *simple*  $m$ -group as one whose associated group has no subgroup other than the identity invariant under the  $m$ -group. It follows that for  $n > 2$  only alternating groups corresponding to first order  $\delta$ -subgroups can be simple. For in any other case,  $(+1, +1, \dots, +1)$ , the identity of the associated  $\epsilon$ -subgroup, is a subgroup thereof invariant under the  $\delta$ -subgroup. Hence the elements of the associated group of the alternating group with  $\epsilon$ -sequence  $(+1, +1, \dots, +1)$  then constitute a subgroup of the associated group invariant under the alternating group. We now proceed to show, on the strength of the corresponding result for ordinary groups, that when  $n > 4$  every alternating group  $H$  corresponding to a first order  $\delta$ -subgroup is a simple. Since the associated  $\epsilon$ -subgroup has but the sole  $\epsilon$ -sequence  $(+1, +1, \dots, +1)$ , the associated ordinary group  $H_0$  of the alternating group  $H$  consists of all elements  $t = t' t'' \dots t^{(m-1)}$  where  $t^{(i)}$  is any positive substitution on the letters of  $\Gamma_i$ , and is thus the direct product of the ordinary alternating groups  $A_1, A_2, \dots, A_{m-1}$  on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  respectively. Let then  $K_0$  be any subgroup of  $H_0$  invariant under  $H$ . If there could be more than one  $t$  in  $K_0$  with the same components  $t', t'', \dots, t^{(m-2)}$ , then there would be more than one  $t$  in  $K_0$  with each component  $t', t'', \dots, t^{(m-2)}$  the identity. Now these  $t$ 's must constitute a subgroup of  $K_0$ , and this subgroup will be invariant under  $H_0$  as a consequence of the invariance of  $K_0$  under  $H_0$ . The corresponding  $t^{(m-1)}$ 's must then constitute an invariant subgroup of  $A_{m-1}$ , if not  $A_{m-1}$  itself. Under the present supposition the last would be true; for with  $n > 4$ , the alternating group  $A_{m-1}$  is simple. But then  $K_0$  would coincide with  $H_0$ , instead of being a subgroup of  $H_0$ . For,  $K_0$  being invariant under any  $s$  in  $H$ , if we transform the above elements of  $K_0$  by  $s, s^2, \dots, s^{(m-2)}$ , we would have in  $K_0$  every element of  $H_0$  any  $m-2$  of whose components are the identity; and the

products of these elements constitute  $H_0$ . We have therefore proved that an element  $t = t't'' \dots t^{(m-1)}$  of  $K_0$  is uniquely determined by its first  $m-2$  components. If then we transform  $t$  by any element of  $H_0$  the first  $m-2$  of whose components are the identity, the first  $m-2$  components of  $t$ , and hence  $t$  itself, will be unchanged.  $t^{(m-1)}$  is then always an invariant element of  $A_{m-1}$ , and, again with  $n > 4$ , can only be the identity. The same argument would show each component of an element of  $K_0$  to be the identity, so that  $K_0$  is the identity.

We have therefore proved that for  $n > 4$  the  $2^{m-2}$   $m$ -adic alternating groups of degree  $n$  corresponding to first order  $\delta$ -subgroups are simple, the others not. For  $n=4$  no  $m$ -adic alternating group is simple, since we can let  $K_0$  be the direct product of the axial groups on the letters of the several  $\Gamma$ 's. Again, for  $n=3$ , no  $m$ -adic alternating group is simple for any  $m > 2$ . The preceding argument breaks down at the one point where the invariance of  $t^{(m-1)}$  under  $A_{m-1}$  is used to prove  $t^{(m-1)}$  the identity.  $K_0$  may now be the third order group obtained from the simple isomorphism between  $A_1, A_2, \dots, A_{m-1}$  that results when  $A_i$  is transformed into  $A_{i+1}$  by a fixed element  $s$  of  $H$ . Finally, when  $n=2$ , the very first step of our argument breaks down. The  $m$ -adic alternating groups can now be identified with the  $\delta$ -subgroups themselves. The simple  $\delta$ -subgroups are those whose associated  $\epsilon$ -subgroups have no admissible  $\epsilon$ -subgroup for subgroup other than the identity. Hence, in terms of the above general determination of admissible  $\epsilon$ -subgroups, the simple  $\delta$ -subgroups are those whose associated  $\epsilon$ -subgroups have  $(x^{m-1}+1)/\tau(x)$  prime.

**13. Transitive  $m$ -adic substitution groups.** Since an  $m$ -adic substitution group  $G$  can carry the letters of  $\Gamma_i$  only into those of  $\Gamma_{i+1}$ , we are led to define a *transitive  $m$ -adic group*  $G$  as one whose substitutions will carry each letter of each  $\Gamma$  into every letter of the succeeding  $\Gamma$ . Clearly, the  $m$ -adic symmetric group of arbitrary degree  $n$ , and the  $m$ -adic alternating groups of degree  $n > 2$  are then transitive. Our analysis in the next section shows that  $G$  will be transitive if the above condition is true for any one  $\Gamma$ , and indeed for any one letter of a  $\Gamma$ , i.e., if the substitutions of  $G$  carry one letter of one  $\Gamma$  into every letter of the succeeding  $\Gamma$ , the same is true of every letter of every  $\Gamma$ , and  $G$  is transitive.

It is readily proved that the containing 2-group  $G^*$  of a transitive  $m$ -group  $G$  is transitive. In fact, let  $a_{ij}$  and  $a_{(i+k)j'}$  be any two letters of the  $\Gamma$ 's. Considering  $i+k$  reduced modulo  $m-1$ , we may assume  $1 \leq k \leq m-1$ . We need not consider  $k=1$ . For  $k > 1$  let  $r$  be any  $(k-1)$ -ad. It will carry  $a_{ij}$  into some  $a_{(i+k-1)j''}$ . Some  $s$  will carry  $a_{(i+k-1)j''}$  into  $a_{(i+k)j'}$ . Hence the  $k$ -ad  $rs$ , which is a substitution in  $G^*$ , carries  $a_{ij}$  into  $a_{(i+k)j'}$  as required. Conversely, if  $G^*$  is transitive, a substitution in  $G^*$  carrying  $a_{ij}$  into  $a_{(i+1)j'}$  belongs to  $G$ , and hence  $G$  is transitive.

In terms of  $G_0$ , the associated 2-group of  $G$ , we likewise see that  $G$  is transitive when and only when the substitutions of  $G_0$  carry each letter of each  $\Gamma$

into every letter of the same  $\Gamma$ . Recalling our definition of the associated constituent groups  $G'_0, G''_0, \dots, G_0^{(m-1)}$  of  $G$ , we thus have that  $G$  is transitive when and only when its associated constituent groups are transitive. As the latter are conjugate, it follows that  $G$  is transitive if any one of its associated constituent groups is known to be transitive.

Let  $(G_0)_{ij}$  be the subgroup of  $G_0$  which consists of all substitutions in  $G_0$  that carry  $a_{ij}$  into itself. If we expand  $G$  in right cosets as regards  $(G_0)_{ij}$ , the members of each single coset carry  $a_{ij}$  into one and the same letter of  $\Gamma_{i+1}$ . Also, if  $s_1$  and  $s_2$  of  $G$  carry  $a_{ij}$  into the same letter,  $s_1 s_2^{-1}$  will be in  $(G_0)_{ij}$ , so that  $s_1$  and  $s_2$  are in the same coset. Each coset therefore consists of all the substitutions of  $G$  carrying  $a_{ij}$  into the corresponding letter. If then  $G$  is transitive of degree  $n$ , there will be exactly  $n$  such cosets, one for each letter of  $\Gamma_{i+1}$ . Hence, the order of  $(G_0)_{ij}$  is equal to  $g/n$  if  $G$  is a transitive group of order  $g$ , and degree  $n$ . *The order of a transitive  $m$ -adic substitution group is therefore a multiple of its degree.* Furthermore, *the number of substitutions of a transitive  $m$ -adic substitution group that carry any letter  $a_{ij}$  into any letter  $a_{(i+1)k}$  is, for all such pairs of letters, equal to the order of the group divided by its degree.*

Since for  $m > 2$  an  $m$ -adic substitution group cannot carry a letter into itself, we have to turn to the associated group of a transitive  $m$ -adic substitution group for an average number of letters theorem. For this purpose we write the substitutions of the associated group in standard cycle form. Observe first that each associated constituent group  $G_0^{(i)}$  being transitive, and of degree  $n$ , the average number of its letters appearing in its substitutions is  $n-1$ . Fixing our attention on  $G_0^{(0)}$ , we consider the subgroup  $H_0^{(0)}$  of  $G_0$  consisting of all the substitutions of  $G_0$  whose component in  $G_0^{(0)}$  is the identity of  $G_0^{(0)}$ . If we expand  $G_0$  in cosets as regards  $H_0^{(0)}$ , each coset is easily seen to consist of all the substitutions of  $G_0$  which have a fixed component in  $G_0^{(0)}$ . Each substitution of  $G_0^{(0)}$  therefore occurs the same number of times in  $G_0$ . It follows that *the average number of letters of each  $\Gamma_i$  occurring in the substitutions of the associated group of a transitive group of degree  $n$  is  $n-1$ .* This is our strongest result. From it, or from our discussion of  $(G_0)_{ij}$ , we also have that *the average number of all letters appearing in the substitutions of the associated group of a transitive  $m$ -adic substitution group of degree  $n$  is  $(m-1)(n-1)$ .* This may also be seen as follows. Since the containing group  $G^*$  of the transitive  $m$ -adic group  $G$  is transitive, and of degree  $(m-1)n$ , the average number of letters in its substitutions is  $(m-1)n-1$ . The total number of letters in its substitutions is then  $(m-1)g[(m-1)n-1]$ ,  $g$  being the order of  $G$ . Of the  $(m-1)g$  substitutions in  $G^*$ , the  $(m-2)g$  substitutions not in  $G_0$  each has its full complement of  $(m-1)n$  letters. The total number of letters in the substitutions of  $G_0$  is thus the remaining  $(m-1)(n-1)g$  letters, whence the result.

**14. Intransitive  $m$ -adic substitution groups.** Let  $G$  be any  $m$ -adic substitution group on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , and let  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$  be the subclasses of the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  respectively into which  $a_{(m-1)1}$

is carried by the elements, dyads,  $\dots$ ,  $(m-1)$ -ads of  $G$ . If  $s$  is any substitution in  $G$ , then as  $r$  ranges through the  $i$ -ads of  $G$ ,  $rs$  ranges through the  $(i+1)$ -ads of  $G$ . Hence  $s$  transforms the letters of  $\Gamma'_i$  in 1-1 fashion into the letters of  $\Gamma'_{i+1}$  for each  $i$ , and thus determines an  $m$ -adic substitution on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ . Furthermore, if  $a_{ij}$  and  $a_{(i+1)k}$  are any two letters of  $\Gamma'_i$  and  $\Gamma'_{i+1}$  respectively, some  $s$  of  $G$  will carry  $a_{ij}$  into  $a_{(i+1)k}$ . For some  $i$ -ad  $r_1$  of  $G$  carries  $a_{(m-1)1}$  into  $a_{ij}$ , and some  $(i+1)$ -ad  $r_2$  of  $G$  carries  $a_{(m-1)1}$  into  $a_{(i+1)k}$ . Hence element  $r_1^{-1}r_2$  of  $G$  carries  $a_{ij}$  into  $a_{(i+1)k}$ . The  $m$ -adic substitutions on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$  obtained from all the substitutions of  $G$  therefore constitute a transitive  $m$ -adic substitution group on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ . If  $G$  is transitive, this group is identical with  $G$ . If  $G$  is not transitive, we may call this group a *transitive constituent group* of the *intransitive group*  $G$ . In that case, by accounting for all the letters of  $\Gamma_{m-1}$ , we obtain a number of transitive constituent groups of  $G$  such that every substitution in  $G$  is the product of a selection of substitutions from the transitive constituent groups of  $G$ .

This result can also be obtained by analysing the containing group  $G^*$  of  $G$ , whence it also appears that the transitive constituent groups of  $G^*$  are the containing groups of the transitive constituent groups of  $G$ .

The direct product and simple isomorphism methods for obtaining intransitive ordinary groups admit of immediate extension to  $m$ -adic groups. In the latter case, let  $G_1$  and  $G_2$  be the same  $m$ -adic substitution group written on different letters. If the letters of  $G_1$  form the sets  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ , of  $G_2$ ,  $\Gamma''_1, \Gamma''_2, \dots, \Gamma''_{m-1}$ , the products of corresponding substitutions in  $G_1$  and  $G_2$  will be  $m$ -adic substitutions on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , where  $\Gamma_i$  consists of all the letters of  $\Gamma'_i$  and  $\Gamma''_i$ . Clearly, an  $m$ -adic substitution group is thus formed simply isomorphic with  $G_1$  and  $G_2$ , but of twice their degree. Similarly for any number of groups obtained by writing a given  $m$ -adic substitution group on different letters.

As for the direct product method, let  $H_1$  and  $H_2$  be  $m$ -adic substitution groups on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$  and  $\Gamma''_1, \Gamma''_2, \dots, \Gamma''_{m-1}$  with all letters distinct. As before, form  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ . Then, if  $s'_i$  and  $s''_i$  be any two substitutions in  $H_1$  and  $H_2$  respectively,  $s'_i s''_i$  will be an  $m$ -adic substitution on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ . The set of all such products clearly constitutes an  $m$ -adic substitution group  $G$  of order equal to the product of the orders of  $H_1$  and  $H_2$ , and degree equal to the sum of their degrees. When  $m=2$ , the existence of an identical element, coupled with the ambiguity of the cycle notation, allows us to consider  $H_1$  and  $H_2$  subgroups of  $G$  which can then be said to be generated by  $H_1$  and  $H_2$ . When  $m>2$  this is no longer possible. We shall therefore refrain from calling  $G$  the direct product of  $H_1$  and  $H_2$ , reserving that phrase for a more special concept found useful in the sequel<sup>(51)</sup>.

#### 15. Substitutions which are commutative with each of the substitutions

<sup>(51)</sup> In fact, while  $G$  is an  $m$ -adic substitution group, the  $m$ -group "generated" by  $H_1$  and  $H_2$  is, for  $m>2$ , a hybrid sort of an affair of order  $m-1$  times the order of  $G$ . On the other hand,

**of a transitive  $m$ -adic substitution group.** Recalling that the order of a transitive  $m$ -adic substitution group is a multiple of its degree, we may most briefly define a *regular  $m$ -adic substitution group* as a transitive  $m$ -adic substitution group whose order is equal to its degree. In view of the corresponding general result for transitive groups, this is equivalent to defining a regular  $m$ -adic substitution group as an  $m$ -adic substitution group, which, for any pair of letters in consecutive  $\Gamma$ 's, has one and only one substitution carrying the first letter into the second. Other transitive group results, coupled with the order criterion of regularity, show that an  $m$ -adic substitution group is regular if and only if its containing group is regular; also, if and only if its associated constituent groups are regular. The orders of the associated group, and the associated constituent groups, then being the same, it also follows that a regular  $m$ -adic substitution group is a transitive group whose associated group has no substitutions other than the identity omitting a letter. Regular  $m$ -adic substitution groups play the same role in polyadic as in ordinary group theory, since we later show that every finite abstract  $m$ -adic group can be represented as a regular  $m$ -adic substitution group.

According to a theorem of Jordan, the substitutions on the letters of a regular group commutative with each of its substitutions constitute a group conjugate to the regular group and known as its conjoint. We extend this theorem to a regular  $m$ -adic substitution group  $G$  by directly applying it to the containing group  $G^*$ , which is known to be regular. In fact, since  $G^*$  is generated by  $G$ , the ordinary substitutions on the letters of  $G$  commutative with each of its substitutions are the same as those commutative with each of the substitutions of  $G^*$ . Hence, to find the  $m$ -adic substitutions on the letters of  $G$  commutative with each of its substitutions we need merely pick out those substitutions in the conjoint of  $G^*$  which are  $m$ -adic substitutions.

To do this we must re-examine the standard proof of Jordan's theorem. In this proof the letters on which the given regular group is written are replaced by the symbols  $s_i$  used for the substitutions in the group. Then, in the simple isomorphism established between the group and its conjoint, the  $j$ th substitution of the given group in its new form replaces each symbol  $s_i$  by the symbol for the substitution  $s_i s_j$ , while the corresponding substitution in the conjoint replaces each  $s_i$  by  $s_j s_i$ . Finally, it is shown that the given group is transformed into its simply isomorphic conjoint by the substitution which carries the second letter of each substitution of the given group into the second letter of the corresponding substitution of the conjoint when all the substitutions<sup>(62)</sup> are written in cycle form with the same first letter.

---

if either  $H_1$  or  $H_2$  has a first order element,  $G$  will contain a subgroup  $H_2'$  or  $H_1'$  simply isomorphic to  $H_2$  or  $H_1$  respectively; and if both  $H_1$  and  $H_2$  possess an invariant first order element, the corresponding  $H_2'$  and  $H_1'$  will generate  $G$ , and  $G$  will then be the direct product of  $H_1'$  and  $H_2'$  in the sense later defined (§25).

(62) All except the identity, that is. Likewise later in the proof.



For our purpose it suffices to determine the nature of this substitution in the case of the regular group  $G^*$ . With  $G$  a regular  $m$ -adic substitution group on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , the substitutions of  $G^*$  can be grouped into corresponding classes  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$  according as they are elements, dyads,  $\dots$ ,  $(m-1)$ -ads of  $G$ . When  $G^*$  is rewritten in accordance with the proof of Jordan's theorem,  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$  take the place of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ . In fact, if  $s_i$  is a  $k$ -ad in  $G^*$ ,  $s_j$  an  $l$ -ad,  $s_i s_j$  is a  $(k+l)$ -ad. Hence, in the above description applied to  $G^*$ , if  $s_j$  is an  $l$ -ad, the  $j$ th substitution of  $G^*$  transforms each  $\Gamma'_k$  into  $\Gamma'_{k+l}$ , and hence is an  $l$ -ad on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ . But the same reasoning shows the corresponding substitution in the conjoint of  $G^*$  also to be an  $l$ -ad on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m-1}$ . Or, returning to  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , we have that in the simple isomorphism between  $G^*$  and its conjoint the correspondant of an  $i$ -ad in  $G^*$  is an  $i$ -ad. If then we write the substitutions of  $G^*$  and its conjoint with the same first letter, say  $a_{11}$ , if the corresponding substitutions in  $G^*$  and its conjoint are both  $i$ -ads, their second letters will both be in  $\Gamma_{i+1}$ . Hence, the substitution which transforms  $G^*$  into its conjoint transforms each  $\Gamma$  into itself, and consequently is an  $(m-1)$ -ad on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ .

Our result now immediately follows. The  $(m-1)$ -ad will transform only the  $m$ -adic substitutions of  $G^*$  into  $m$ -adic substitutions. Hence the  $m$ -adic substitutions in the conjoint of  $G^*$  are the transforms of the  $m$ -adic substitutions in  $G^*$  by the  $(m-1)$ -ad, i.e., the transforms of the substitutions in  $G$ . Since the transform of an  $m$ -adic group is a simply isomorphic  $m$ -adic group, we thus have the following extension of Jordan's theorem. *The  $m$ -adic substitutions on the sequence of  $\Gamma$ 's of a regular  $m$ -adic substitution group commutative with all the substitutions of the group constitute a regular  $m$ -adic substitution group of the same order, and this group is the transform of the given group by an  $(m-1)$ -ad of  $m$ -adic substitutions.* Clearly, the relationship between the two groups is a reciprocal one, and we may call each the *conjoint* of the other. Either directly, or as a consequence of a later general result on transforms, it may be verified that each group can be transformed into the other by an  $m$ -adic substitution, and hence, according to any  $m$ -adic definition, are conjugate.

We further have, as a result of the above discussion, that every ordinary substitution on the letters of a regular  $m$ -adic substitution group of degree  $n$  commutative with all the substitutions of the group are polyads of  $m$ -adic substitutions, there being  $n$  such  $i$ -ads for every  $i$ . Together they of course constitute the conjoint of the containing group of the given group; and this conjoint is now seen to be the containing group of the conjoint of the given group.

In passing from regular  $m$ -adic substitution groups to arbitrary transitive  $m$ -adic substitution groups for the purpose of extending Kuhn's theorem to  $m$ -adic groups, we shall adopt the viewpoint of the last paragraph, and seek



all substitutions on the letters of the transitive  $m$ -adic group commutative with each of its substitutions; for now  $m$ -adic substitutions of this kind will exist only if the given group satisfies a special condition. If  $G$  is a transitive  $m$ -adic substitution group,  $G^*$  is transitive. Again, the substitutions on the letters of  $G$  commutative with every substitution in  $G$  are the substitutions on the letters of  $G^*$  commutative with each of its substitutions, and hence can be found by applying Kuhn's theorem to  $G^*$ . As before, we assume all substitutions written in cycle form.

According to Kuhn's generalization of Jordan's theorem the number of substitutions on the letters of  $G^*$  commutative with each of its substitutions is the same as the number of letters omitted in all substitutions of  $G^*$  which omit a given letter. Actually, such substitutions will be in  $G_0$ , the associated group of  $G$ . Let  $\{a_{ij}\}$  designate the set of letters omitted by all substitutions of  $G^*$  that omit  $a_{ij}$ . Since  $G^*$  is transitive, it follows that if  $a_{i_1j_1}$  is in  $\{a_{ij}\}$ , then  $\{a_{i_1j_1}\} = \{a_{ij}\}$ , and a substitution  $r$  of  $G^*$ , carrying  $a_{ij}$  into  $a_{i_1j_1}$ , carries the set of letters  $\{a_{ij}\}$  into itself. But  $r$  carries all the letters of  $\Gamma_i$  into all those of  $\Gamma_{i_1}$ , and hence all the letters of  $\{a_{ij}\}$  that are in  $\Gamma_i$  into all the letters of  $\{a_{ij}\}$  that are in  $\Gamma_{i_1}$ . Hence, if there are  $\alpha$  letters of  $\{a_{ij}\}$  in one  $\Gamma$ , there are  $\alpha$  letters of  $\{a_{ij}\}$  in every  $\Gamma$  that has at least one of them. Now with the  $\Gamma$ 's arranged in a cycle, let  $\delta$  be the least difference between the subscripts of consecutive  $\Gamma$ 's that have letters of  $\{a_{ij}\}$ . Then some  $\delta$ -ad  $r$  in  $G^*$  will transform the set  $\{a_{ij}\}$  into itself. As  $r$  will then carry the letters of  $\{a_{ij}\}$  which are in any  $\Gamma_i$  into letters of  $\{a_{ij}\}$  which are in  $\Gamma_{i+\delta}$ , it follows that the  $\Gamma$ 's having letters in  $\{a_{ij}\}$  have subscripts which are in arithmetic progression, with the common difference, indeed, a divisor of  $m-1$ . Finally, the known properties of transitive groups show the different sets  $\{a_{ij}\}$  to be mutually exclusive, and transformable into each other by the substitutions of  $G^*$ . It follows that the numbers  $\alpha$  and  $\delta$  are the same for all such sets; and since together they exhaust the letters of  $G^*$ , that  $\alpha$  is a divisor of  $n$ . Hence the following result. *If  $G$  is a transitive  $m$ -adic substitution group of degree  $n$ , then the number of letters omitted by all substitutions of the containing group  $G^*$  that omit a given letter is of the form  $\kappa\alpha$ , where  $\kappa$  is a divisor of  $m-1$ ,  $\alpha$  a divisor of  $n$ ; furthermore, there are  $\alpha$  of these letters in every  $\Gamma$  that has at least one, the subscripts of these  $\Gamma$ 's forming an arithmetic progression.*

According to the proof of Kuhn's theorem the resulting  $\kappa\alpha$  substitutions on the letters of  $G^*$  commutative with each of its substitutions are obtained as follows. Let  $H_{11}$  be the subgroup of  $G^*$  composed of the substitutions of  $G^*$  which leave the set of letters  $\{a_{11}\}$  unchanged, and let  $C_{11}$  be the conjoint of the regular group  $K_{11}$ , on the letters  $\{a_{11}\}$ , formed by the components on those letters of the substitutions in  $H_{11}$ . For each substitution in  $C_{11}$ , form the product of all the distinct transforms of that substitution under  $G^*$ . These products are the  $\kappa\alpha$  substitutions on the letters of  $G^*$  commutative with each of its substitutions. According to our distribution result

for the set of letters  $\{a_{11}\}$ , there are  $\alpha$  of these letters in each of the  $\kappa$   $\Gamma$ 's,  $\Gamma_1, \Gamma_{1+\delta}, \dots, \Gamma_{1+(\kappa-1)\delta}$ , where  $\kappa\delta = m-1$ . Clearly, the substitutions of  $H_{11}$  can only be  $i$ -ads with  $i = \delta, 2\delta, \dots, \kappa\delta$ . Since  $K_{11}$  is regular on the letters  $\{a_{11}\}$ , it will have, for each of the above  $i$ 's,  $\alpha$  substitutions which are components of  $i$ -ads in  $H_{11}$ . Our proof of the extended Jordan theorem applies sufficiently to the relationship between  $K_{11}$  and its conjoint  $C_{11}$  to show that  $C_{11}$  also consists of  $\alpha$  "components of  $i$ -ads" for each  $i = \delta, 2\delta, \dots, \kappa\delta$ . Now when a substitution of  $C_{11}$  is transformed by the substitutions of  $G^*$ , the set of letters  $\{a_{11}\}$  will go over into all the mutually exclusive distinct sets  $\{a_{ij}\}$ , there being one and only one distinct transform of the substitution of  $C_{11}$  for each set  $\{a_{ij}\}$ . If the substitution in question is a component of an  $i$ -ad, each transform will also be a component of an  $i$ -ad. As the sets  $\{a_{ij}\}$  are mutually exclusive, and exhaust the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , the product of these transforms will exactly constitute an  $i$ -ad. We thus have the following extension of Kuhn's theorem. *The only substitutions on the letters of the  $\Gamma$ 's of a transitive  $m$ -adic substitution group commutative with each of its substitutions are polyads of  $m$ -adic substitutions on the same sequence of  $\Gamma$ 's; in the notation of the distribution theorem, if  $\delta = (m-1)/\kappa$ , these polyads can only be  $i$ -ads with  $i = \delta, 2\delta, \dots, \kappa\delta$ , there being exactly  $\alpha$  such  $i$ -ads for each admissible  $i$ .*

In particular, if we restrict our attention to  $m$ -adic substitutions, we have the following result. *The necessary and sufficient condition that there be at least one  $m$ -adic substitution on the sequence of  $\Gamma$ 's of a transitive  $m$ -adic substitution group commutative with each of its substitutions is that the subgroup of the associated group consisting of all its substitutions omitting a given letter in one  $\Gamma$  omits a fixed letter in the following  $\Gamma$ ; if then that subgroup omits exactly  $\alpha$  letters from one  $\Gamma$ , it will omit  $\alpha$  letters from every  $\Gamma$ , and there will be exactly  $\alpha$  such  $m$ -adic substitutions.*

**16. Holomorphs of a regular  $m$ -adic substitution group.** The concept of holomorph of a regular group admits both of an immediate extension to regular  $m$ -adic substitution groups, as well as of a further generalization peculiar to polyadic theory. For the immediate extension, let  $G$  be a regular  $m$ -adic substitution group of order  $n$  on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ . Then all the  $m$ -adic substitutions on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  which transform  $G$  into itself constitute an  $m$ -adic substitution group of degree  $n$  which we shall call the *principal holomorph* of  $G$ . Clearly, the principal holomorph of  $G$  not only contains  $G$ , but also the conjoint of  $G$ . Since the transforms of commutative substitutions are commutative, it follows that the principal holomorph of  $G$  is in fact also the principal holomorph of its conjoint.

If  $K$  is the principal holomorph of  $G$ , then  $(K_0)_{11}$ , the subgroup of the associated group  $K_0$  of  $K$  consisting of all the substitutions of  $K_0$  omitting  $a_{11}$ , may be identified as the *group of isomorphisms* of  $G$ . That is,  $(K_0)_{11}$  transforms  $G$  into all of its possible automorphisms, each automorphism being given by but one substitution of  $(K_0)_{11}$ . In fact, the argument used in extending

Jordan's theorem shows that the substitutions of one of two simply isomorphic regular  $m$ -adic substitution groups on the same sequence of  $\Gamma$ 's can be transformed into the corresponding substitutions of the other by an  $(m-1)$ -ad which omits, say,  $a_{11}$ . Hence, every automorphism of  $G$  can be obtained by transforming  $G$  by the substitutions in  $(K_0)_{11}$ . Furthermore, if two distinct substitutions of  $(K_0)_{11}$  yielded the same automorphism of  $G$ , a substitution of  $(K_0)_{11}$  other than the identity would transform each member of  $G$  into itself. But this substitution would have to be in the associated group of the conjoint of  $G$ , and, as this conjoint is regular, the substitution in question, which omits  $a_{11}$ , could only be the identity.

We can now prove, as in the ordinary case, that *the order of the principal holomorph of a regular  $m$ -adic substitution group is equal to the product of the order of the group and the order of its group of isomorphisms*. In fact, if  $\bar{G}$  is the conjoint of the regular group  $G$ ,  $K$  its holomorph, by expanding  $K$  in cosets as regards its invariant subgroup  $\bar{G}$ , we see that the substitutions of  $K$  transform  $G$  in  $k/n$  different ways,  $k$  being the order of  $K$ . But  $K$  as well as  $(K_0)_{11}$  must transform  $G$  into all of its possible automorphisms. For if  $s$  is in  $K$ ,  $t$  in  $(K_0)_{11}$ , as  $t$  runs through  $(K_0)_{11}$  giving all the automorphisms of  $G$ ,  $ts$  in  $K$  yields an equal number of automorphisms of  $G$ . Hence the order of  $(K_0)_{11}$  is  $k/n$ , whence the above result.

To illustrate this result, consider the cyclic triadic group of degree and order two generated by the triadic substitution  $s_1 = (a_{11}a_{21}a_{12}a_{22})$  given in cycle form. The letters of  $\Gamma_1$  are  $a_{11}, a_{12}$ , of  $\Gamma_2$ ,  $a_{21}, a_{22}$ . The sole other triadic substitution on  $\Gamma_1, \Gamma_2$  generated by  $s_1$  is  $s_2 = (a_{11}a_{22}a_{12}a_{21})$ , so that the group is seen to be regular.  $s_2$  also is a generator of the group, whence it follows that the group admits exactly two automorphisms. Hence the order of its principal holomorph is four. We find directly that  $s_3 = (a_{11}a_{21})(a_{12}a_{22})$  and  $s_4 = (a_{11}a_{22})(a_{12}a_{21})$  interchange  $s_1$  and  $s_2$ , so that the principal holomorph consists of  $s_1, s_2, s_3, s_4$ . It is actually the entire triadic symmetric group of degree two. This example serves to answer the question whether some subgroup of the principal holomorph itself, instead of its associated ordinary group, can be identified with the group of isomorphisms of the given group. The answer in the present instance is no. For such a subgroup would have to possess as element  $s_1$  or  $s_2$  to yield the identical automorphism, but would then have for elements both  $s_1$  and  $s_2$ , each yielding that one automorphism.

The immediate extension of the concept of complete group to  $m$ -adic groups turns out to be rather trivial. Defining an  $m$ -adic group  $G$  to be *complete in the narrow sense* if its own elements transform it in 1-1 fashion into all of its possible automorphisms, we obtain the following result. *An  $m$ -group is complete in the narrow sense when and only when it is reducible to a complete ordinary group*. In fact, its sole element yielding the identical automorphism must be of first order, and invariant under the group—hence the reducibility.

The rest of the theorem follows from the easily demonstrated facts that if  $G$  is reducible to  $G'$ , every automorphism of one group is also an automorphism of the other, while the automorphisms induced by any element of either is the same for both. Since  $G$  can have but one invariant element, it also follows that *the net of derived groups of an  $m$ -adic group complete in the narrow sense consists of a single complete 2-group, and its extensions, which are then also complete in the narrow sense.* If  $G$  is regular, and complete in the narrow sense, we may use its elements as multipliers in the expansion of  $K$  in cosets as regards  $\bar{G}$ . We shall express this fact by saying that *the principal holomorph of an  $m$ -group complete in the narrow sense is the direct product of the group and its conjoint.* A precise abstract definition of this rather narrow concept of direct product will be given in §25.

We do not obtain a less restrictive concept of completeness by asking that the elements of  $G_0$  transform  $G$  into all of its possible automorphisms in 1-1 fashion; for the coset theorem shows that  $G$  and  $G_0$  transform  $G$  according to the same number of distinct automorphisms. We therefore define  $G$  to be *complete in the wide sense* if the elements of its abstract containing group  $G^*$  transform it in 1-1 fashion according to all of its possible automorphisms. Since only the identity of  $G^*$  is now invariant under  $G$ , it follows that an  $m$ -group complete in the wide sense is irreducible. If this  $m$ -group  $G$  is of order  $g$ , and is expressed as a regular  $m$ -adic substitution group, the order of its principal holomorph  $K$  will be  $(m-1)g^2$ . We now turn to the containing groups for a direct product theorem, and easily find that *the containing group of the principal holomorph of an  $m$ -group complete in the wide sense is the direct product of the containing groups of the group and its conjoint.*

Actually, a type of completeness can be defined for each divisor  $k$  of  $m-1$ , an  $m$ -group  $G$  being said to be complete in the  $k$ -sense if it admits some containing group of index  $k$  whose elements yield in 1-1 fashion all the automorphisms of  $G$ . We then have that an  $m$ -group is complete in the  $k$ -sense when and only when it is reducible to a  $(k+1)$ -group complete in the wide sense. Furthermore, the net of derived groups of an  $m$ -group complete in the  $k$ -sense consists of a single  $(k+1)$ -group complete in the wide sense, and its extensions. With  $G$  written as a regular  $m$ -adic substitution group, its principal holomorph will of course be of order  $kg^2$ . But there does not then seem to be a direct product theorem in terms of groups.

We turn now to the purely  $m$ -adic generalization of holomorph. In ordinary group theory, due to the presence of the identity, if all of the elements of a group  $H$  transform a group  $G$  into one and the same group, that group must be  $G$  itself. Hence if  $G$  is a regular substitution group,  $H$  a substitution group on the letters of  $G$ ,  $H$  will be the holomorph of  $G$ , or a subgroup thereof. This need not be so for  $m$ -adic groups with  $m > 2$ . Let then  $G$  be a regular  $m$ -adic substitution group with  $m > 2$ ,  $H$  an  $m$ -adic substitution group on the

sequence of  $\Gamma$ 's of  $G$  such that all of the substitutions of  $H$  transform  $G$  into one and the same group  $G''^{(u)}$ . It follows that all of the substitutions of  $H$  transform  $G''$  into one and the same group  $G'''$ , and so on. Since the  $m$ -ads of  $H$  must transform  $G$  as do its elements, it follows that there will be a cycle of  $\mu-1$  distinct, though not necessarily mutually exclusive,  $m$ -adic groups  $(G', G'', \dots, G^{(\mu-1)})$ , such that  $G' = G$ ,  $\mu-1$  is a divisor of  $m-1$ , and all the elements of  $H$  transform each  $G$  into the cyclically following  $G$ . Now all the  $m$ -adic substitutions on the sequence of  $\Gamma$ 's of  $G$  which transform  $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(\mu-1)} \rightarrow G$  constitute an  $m$ -adic substitution group  $K$  containing  $H$ . We shall then call  $K$  a *holomorph* of  $G$ , and the *holomorph* of the cycle  $(G', G'', \dots, G^{(\mu-1)})$ . When  $\mu-1=1$ ,  $K$  becomes the principal holomorph of  $G$ .

Given the regular  $G$ , an  $m$ -adic substitution  $s$  on the sequence of  $\Gamma$ 's of  $G$  will be said to be *holomorphic* if it belongs to some holomorph of  $G$ . We then readily see that *the necessary and sufficient condition that  $s$  be holomorphic is that  $s^{m-1}$  is in the associated group of the principal holomorph of  $G$* . For that associated group consists of all the  $(m-1)$ -ads on the sequence of  $\Gamma$ 's of  $G$  which transform  $G$  into itself. The necessity of the condition then follows from the fact that  $s^m$  must transform  $G$  into the same group that  $s$  does, the sufficiency from the fact that all the elements of the cyclic  $m$ -group generated by  $s$  will then transform  $G$  into one and the same group. In particular, the  $(n!)^{m-2}$  first order substitutions of degree  $n$  are all holomorphic for the regular  $G$  of degree  $n$ . Hence, when the order of the principal holomorph of  $G$  is less than  $(n!)^{m-2}$ , as must be so, for example, in the case of cyclic  $m$ -groups of order greater than three, we are assured of the existence of a holomorph other than the principal holomorph. Clearly, any element  $s$  of a holomorph of  $G$  determines the corresponding cycle  $(G', G'', \dots, G^{(\mu-1)})$ , and hence the holomorph. It follows that all the holomorphs of a given  $G$  are mutually exclusive.

Our next result shows that the order of any holomorph of  $G$  is no greater than that of the principal holomorph of  $G$ . In fact, let  $K', K'', \dots, K^{(\mu-1)}$  be the principal holomorphs of  $G', G'', \dots, G^{(\mu-1)}$ ,  $K$  the holomorph of the cycle  $(G', G'', \dots, G^{(\mu-1)})$ . By writing an element  $t$  of  $K_0$  as the product of  $m-1$  elements of  $K$ , we see that  $t$  must leave each  $G^{(i)}$  invariant, and hence be in each  $K_0^{(i)}$ . Conversely, if  $t$  is in each  $K_0^{(i)}$ , it will transform each  $G^{(i)}$  into itself. If then  $s$  is in  $K$ ,  $ts$  will also be in  $K$ , so that  $t$  must be in  $K_0$ . That is, *the associated group of the holomorph of  $(G', G'', \dots, G^{(\mu-1)})$  is the logical product of the associated groups of the principal holomorphs of  $G', G'', \dots, G^{(\mu-1)}$* . It is readily verified that an  $s$  in  $K$  actually transforms  $K' \rightarrow K'' \rightarrow \dots \rightarrow K^{(\mu-1)} \rightarrow K'$ , and hence also  $K'_0 \rightarrow K''_0 \rightarrow \dots \rightarrow K^{(\mu-1)}_0 \rightarrow K'_0$ . Hence, a subgroup of  $K'_0$  invariant under  $s$  must be contained in  $K_0$ . We thus have, in terms of  $G$  alone, *the associated group of the holomorph of  $G$  corresponds*

<sup>(u)</sup> The reader will note the marked analogy with Corral's concept of a function pertaining to a brigade, that is, one carried into the same function by all the substitutions of the brigade.



ing to a holomorph  $s$  is the largest group or subgroup of the associated group of the principal holomorph of  $G$  invariant under  $s$ .

On turning to an order theorem for these  $m$ -adic holomorphs, we observe first that the holomorph of a cycle  $(G', G'', \dots, G^{(\mu-1)})$  is also the holomorph of the "conjoint cycle"  $(\bar{G}', \bar{G}'', \dots, \bar{G}^{(\mu-1)})$ . For, inasmuch as transforms of commutative substitutions are commutative, transforms of conjoint regular groups are conjoint. Hence, if  $s$  transforms  $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(\mu-1)} \rightarrow G'$ , it must transform  $\bar{G}' \rightarrow \bar{G}'' \rightarrow \dots \rightarrow \bar{G}^{(\mu-1)} \rightarrow \bar{G}'$ , and conversely. Now let  $K$  be the holomorph of the cycle  $(G', G'', \dots, G^{(\mu-1)})$ . Each  $s$  in  $K$  transforms in 1-1 fashion the elements of  $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(\mu-1)} \rightarrow G'$ , and hence determines a  $\mu$ -adic substitution on the  $\mu-1$  classes of elements  $G', G'', \dots, G^{(\mu-1)}$  which may be termed a  $\mu$ -adic automorphism of the cycle  $(G', G'', \dots, G^{(\mu-1)})$ . The class of all such  $\mu$ -adic substitutions on  $G', G'', \dots, G^{(\mu-1)}$  obtained through substitutions in  $K$  clearly constitutes an  $m$ -adic group which we shall term the *restricted  $m$ -adic group of isomorphisms* of the cycle  $(G', G'', \dots, G^{(\mu-1)})$ , restricted, both by the possible narrowness of  $K$ , and by the fact that while an  $m$ -adic substitution will transform any one  $G^{(i)}$  into  $G^{(i+1)}$  according to any simple isomorphism, it need not be able to do this arbitrarily and simultaneously for each  $i$ . Now  $s_1$  and  $s_2$  of  $K$  will yield the same  $\mu$ -adic automorphism of the cycle  $(G', G'', \dots, G^{(\mu-1)})$  when and only when the  $(m-1)$ -ad  $s_2 s_1^{-1}$  transforms each element of each  $G^{(i)}$  into itself, and hence, when and only when  $s_2 s_1^{-1}$  is in  $\bar{G}_0 = \bar{G}_0' \bar{G}_0'' \dots \bar{G}_0^{(\mu-1)}$ <sup>(4)</sup>. Note that  $K_0$  consists of all  $(m-1)$ -ads which transform each  $G^{(i)}$  into itself, and hence has  $\bar{G}_0$  for subgroup, one, indeed, invariant under  $K$ . By expanding  $K$  in cosets as regards  $\bar{G}_0$ , we then obtain the following result. *The order of the holomorph of  $(G', G'', \dots, G^{(\mu-1)})$  is the product of the order of the crosscut of the associated groups of the conjoints of  $G', G'', \dots, G^{(\mu-1)}$  and the order of the restricted  $m$ -adic group of isomorphisms of  $(G', G'', \dots, G^{(\mu-1)})$ .*

This result is weaker than the result for the principal holomorph of  $G$  in two ways. On the one hand, the order of  $\bar{G}_0$  replaces the order of  $G$  itself. More significantly, in the case of the principal holomorph, we identified  $(K_0)_{11}$  with the group of isomorphisms of  $G$ . Note that there both  $K$  and  $K_0$  yielded every possible automorphism of  $G$ . In the present case  $K_0$  transforms each  $G^{(i)}$  into itself, the distinct transformations being  $(\mu-1)$ -ads of  $\mu$ -adic substitutions on  $G', G'', \dots, G^{(\mu-1)}$ , and constituting the associated group of the restricted  $m$ -adic group of isomorphisms of the cycle  $(G', G'', \dots, G^{(\mu-1)})$ . If then we ask whether  $(K_0)_{11}$  can be identified with this associated restricted group of isomorphisms, we find that while no two members of  $(K_0)_{11}$  can transform the  $G^{(i)}$ 's in the same way, for  $(K_0)_{11}$  to transform the  $G$ 's in every way that  $K_0$  does, it is necessary and sufficient that  $K_0$  carry  $a_{11}$  into no other letters than does its subgroups  $\bar{G}_0$ . We have not succeeded in answering the

<sup>(4)</sup> Product here is logical product.



question thus posed; and hence, whether  $(K_0)_{11}$ , or any other subgroup of  $K_0$ , can be identified as the associated restricted group of isomorphisms of the cycle  $(G', G'', \dots, G^{(\mu-1)})$  remains one of our unsolved problems<sup>(55)</sup>.

17. *m*-adic groups of  $\mu$ -adic substitutions. The present extension of the concept of *m*-adic substitution group is indispensable for a self-contained theory of primitivity, our next topic. This extension has the advantage of including *m*-adic groups of ordinary substitutions in its scope. However, the fact that any abstract *m*-adic group can be represented as a regular *m*-adic substitution group is perhaps sufficient reason for our restricting the explicit study of this wider class of substitution groups to the next section.

Given a cycle of  $\mu-1$  equivalent classes  $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$ , not only will the product of  $\mu$   $\mu$ -adic substitutions on these  $\Gamma$ 's be a substitution of the same kind, but also the product of any *m* such substitutions, provided *m* is in the form  $k(\mu-1)+1$ . We are thus led to the concept of an *m*-adic group of  $\mu$ -adic substitutions, or  $(m, \mu)$  substitution group, with *m* and  $\mu$  subject to the sole condition that  $\mu-1$  be a divisor of *m*-1. We have already met this concept in the last section where the corresponding  $\Gamma$ 's,  $G', G'', \dots, G^{(\mu-1)}$ , while distinct, were probably not necessarily mutually exclusive<sup>(56)</sup>. In what follows, for simplicity, as in our previous development, we shall assume the  $\Gamma$ 's to be mutually exclusive.

It is not difficult to review our previous work to see how much goes over to  $(m, \mu)$  substitution groups. The chief failure turns out to be the extension of Jordan's theorem on regular groups. Particular mention must be made of the structure of the containing group of an  $(m, \mu)$  group *G*. Letting, for simplicity,  $G^*$  symbolize the containing ordinary group of *G* generated by the elements of *G*,  $G^*$  will now be of some index *k* which is a divisor of  $m-1$  and a multiple of  $\mu-1$ . We must now distinguish between *i*-ads of *G* and *i*-ads in  $G^*$ , the former being the products of any *i* substitutions in *G*, the latter all products in  $G^*$  of  $\mu$ -adic substitutions. In particular, there will be  $k/(\mu-1)$  cosets in  $G^*$  consisting of  $(\mu-1)$ -ads, one and only one of these cosets being  $G_0$ .

In connection with the next section, the extension of the concept of transitivity to  $(m, \mu)$  substitution groups is of most importance. Actually, our definition of transitivity as applied to *m*-adic substitution groups can be restated

<sup>(55)</sup> The above theory of *m*-adic holomorphs can be paraphrased for ordinary groups. Thus, if *s* is a substitution on the letters of an ordinary regular group  $G'$ , but not in the holomorph of  $G'$ , and if  $s^{m-1}$  is the first positive power of *s* in the holomorph of  $G'$ , then a cycle of regular groups  $G', G'', \dots, G^{(m-1)}$  is determined such that, under *s*,  $G' \rightarrow G'' \rightarrow \dots \rightarrow G^{(m-1)} \rightarrow G'$ . The set of all substitutions on the letters of  $G'$  thus transforming this now given cycle of  $G$ 's will then constitute an *m*-adic group of ordinary substitutions, which may then be called an *m*-adic holomorph of *G*. The above theory, in somewhat simpler form, will then go over.

<sup>(56)</sup> On the other hand, the most general possibility is still not there illustrated; for a given element is carried into a single element independently of the  $G^{(i)}$  of which it is an element. Note that our last footnote further introduced an  $(m, 2)$  substitution group as *m*-adic holomorph of an ordinary regular group.

verbatim for  $(m, \mu)$  substitution groups. It is then readily verified that all of the results of §13 go over, with the possible replacement of  $m$  by  $\mu$ , with one exception. And that is that the transitivity of  $G^*$  no longer assures the transitivity of  $G^{(57)}$ .

**18. Primitive and imprimitive  $(m, \mu)$  substitution groups.** The distinct sets  $\{a_{ij}\}$  of §15 are transformed as units under all the substitutions of the containing group  $G^*$  of the transitive  $m$ -adic substitution group  $G$ , and hence under the substitutions of  $G$ . We recall that each set  $\{a_{ij}\}$  had  $\alpha$  letters in each of  $\kappa$   $\Gamma$ 's whose subscripts formed an arithmetic progression,  $\alpha$  being a divisor of  $n$ , the degree of  $G$ ,  $\kappa$  of  $m-1$ . Let  $\nu = n/\alpha$ ,  $m'-1 = (m-1)/\kappa$ . Each set  $\{a_{ij}\}$  then has letters in one and only one of the first  $(m'-1)$   $\Gamma$ 's, there being  $\nu$  sets for each such  $\Gamma$ . The  $(m'-1)\nu$  distinct sets  $\{a_{ij}\}$  thus fall into  $m'-1$  mutually exclusive classes  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m'-1}$  of  $\nu$  members each. As any  $m$ -adic substitution  $s$  in  $G$  transforms each  $\Gamma_i$  into  $\Gamma_{i+1}$ , it will in 1-1 fashion transform the members of  $\Gamma'_1 \rightarrow \Gamma'_2, \Gamma'_2 \rightarrow \Gamma'_3, \dots, \Gamma'_{m'-1} \rightarrow \Gamma'_1$ , and so define an  $m'$ -adic substitution on  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{m'-1}$ . The totality of these  $m'$ -adic substitutions will then constitute an  $(m, m')$  substitution group  $G'$  of degree  $\nu$ . As  $G$  is transitive, it follows that  $G'$  is transitive, there being, as in the ordinary case, a  $(1, N)$  isomorphism between  $G'$  and  $G$ . With a restriction to be noted later,  $G$  will be said to be imprimitive with systems of imprimitivity  $\{a_{ij}\}$  whenever  $1 < (m'-1)\nu < (m-1)n$ , this however, as in the ordinary case, being but an example of the general concept of imprimitivity.

We thus see that even if we start with transitive  $m$ -adic substitution groups, i.e.,  $(m, m)$  groups, we are led to  $(m, \mu)$  groups. This extension is however sufficient for our purpose. For if we start with a transitive  $(m, \mu)$  substitution group  $G$ , and define the sets  $\{a_{ij}\}$  as before, we obtain, by the same argument, an analogous distribution theorem, and then, as above, a transitive  $(m, \mu')$  substitution group  $G'$  with  $\mu'-1$  a divisor of  $\mu-1$ .

In general, then, let  $G$  be a transitive  $(m, \mu)$  substitution group of degree  $n$  on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$ , with, of course,  $\mu-1$  a divisor of  $m-1$ , and let there be some separation of the  $(\mu-1)n$  letters of the  $\Gamma$ 's into mutually exclusive classes such that these classes are transformed as units under all the substitutions of  $G$ , and hence of  $G^*$ . An entirely analogous argument to the one used in determining the distribution of the letters in the sets  $\{a_{ij}\}$  leads to a corresponding conclusion here. That is, each class consists of the same number  $\kappa\alpha$  of letters, with  $\alpha$  a divisor of  $n$ ,  $\kappa$  of  $\mu-1$ , there being  $\alpha$  letters in each of  $\kappa$   $\Gamma$ 's whose subscripts are in arithmetic progression. As with the  $\{a_{ij}\}$ 's, each such

(57) On the other hand, if  $G^*$  is transitive, we may form a sequence of mutually exclusive  $\Gamma$ 's,  $\Gamma'_1, \Gamma'_2, \dots, \Gamma'_{\mu'-1}$ , such that  $G$  is a transitive  $(m, \mu')$  group on the  $\Gamma''$ 's. Here  $\mu'-1$  is a multiple of  $\mu-1$ , and a divisor of  $m-1$ ; while the  $\Gamma''$ 's are successively subclasses of the  $\Gamma$ 's run through cyclically  $(\mu'-1)/(\mu-1)$  times, and together exhausting the  $\Gamma$ 's. If we call such an  $(m, \mu)$ -group  $G$  semi-transitive, then the main result of §14 goes over for an arbitrary  $(m, \mu)$ -group if we replace the transitive constituent groups by semi-transitive constituent groups.

separation of the letters of the  $\Gamma$ 's leads to a transitive  $(m, \mu')$  substitution group  $G'$  of degree  $\nu$ , where  $\nu = n/\alpha$ ,  $\mu' - 1 = (\mu - 1)/\kappa$ . The numbers  $\alpha$  and  $\kappa$ , of course, depend on the separation in question.

In accordance with the standard definition we would then define  $G$  to be imprimitive if some such separation into mutually exclusive classes is possible with  $1 < (\mu' - 1)\nu < (\mu - 1)n$ , the classes then being the corresponding systems of imprimitivity of  $G$ . This restriction is equivalent to  $\kappa$  and  $\alpha$  not being both one, or simultaneously equal to  $\mu - 1$  and  $n$  respectively. But then  $G$  would always be imprimitive for  $\mu > 2$ , since its substitutions transform the  $\Gamma$ 's themselves as units. We therefore exclude the case  $\alpha = n$ , and thus have the following definition.  *$G$  will be said to be imprimitive if it admits systems of imprimitivity for which  $\alpha < n$ ,  $\kappa$  and  $\alpha$  not both unity; otherwise  $G$  will be said to be primitive.*

The rather artificial restriction  $\alpha < n$  is entirely natural in the case  $\mu = 2$ , and, indeed, we have here the only complete generalizations of the primitivity theorems of ordinary groups.  $G$  is now an  $m$ -adic group of ordinary substitutions on letters which may then be written  $a_1, a_2, \dots, a_n$ . It is easily seen that if  $G$  is transitive, so are  $G^*$  and  $G_0$ , the converse however holding only for  $G_0$ . On the other hand, with  $G$  transitive, if  $G$  is imprimitive, so are  $G^*$  and  $G_0$ , the converse holding only for  $G^*$ . Thus, with  $m = n + 1$ ,  $G$  can be the intransitive group consisting of the single substitution  $(a_1 a_2 \dots a_n)$ , while  $G^*$  is transitive. And the following is an example of a transitive primitive  $G$  for which  $G_0$  is imprimitive. Let  $G_0$  be the transitive imprimitive group of order four:  $1, (a_1 a_2)(a_3 a_4), (a_1 a_3)(a_2 a_4), (a_1 a_4)(a_2 a_3)$ . Then  $s = (a_1 a_2 a_3)$  transforms  $G_0$  into itself, while  $s^3 = 1$  is in  $G_0$ . Hence  $G = G_0 s$  is a transitive tetradic group of ordinary substitutions, and is easily seen to be primitive.

Turning to the ordinary theorems on primitivity, with  $G$  thus a transitive  $(m, 2)$  group, there will be at least one substitution in  $G$  carrying  $a_1$  into itself, and the totality of these substitutions will constitute a subgroup  $G_1$  of  $G$ . We then have the complete analogue of the corresponding theorem for ordinary substitution groups, i.e., *a necessary and sufficient condition that a transitive  $m$ -adic group  $G$  of ordinary substitutions is imprimitive is that  $G_1$  is contained in a larger subgroup of  $G$* . While this can be proved by applying the ordinary theorem to  $G^*$ , the ordinary proof<sup>(88)</sup> can here be directly carried over. Thus, if  $G$  is imprimitive, the substitutions of  $G$  transforming the system of imprimitivity of which  $a_1$  is a member into itself constitute a subgroup  $K$  of  $G$  containing  $G_1$ , and larger than  $G_1$ . Conversely, if  $K$  is a subgroup of the transitive  $G$  containing  $G_1$ , and larger than  $G_1$ , expand  $G$  in right cosets as regards  $K$ . Each coset will consist of the same number  $\alpha > 1$  of right cosets of  $G$  as regards  $G_1$ , and hence will carry  $a_1$  into  $\alpha$  letters, distinct for each coset, and will

<sup>(88)</sup> Rather what the proof of the more general theorem of *Finite Groups*, page 39, would become if given directly for the more special result. We have interchanged the order of the two results.

consist of all the substitutions of  $G$  carrying  $a_1$  into one of those letters. Finally, any substitution  $s$  of  $G$  will transform these mutually exclusive sets of letters as units. For if  $K_{0s_1}$  is the coset carrying  $a_1$  into one of these sets, that set will be transformed by  $s$  as  $a_1$  is by  $K_{0s_1}s$ . But  $K_{0s_1}s$  is the same as  $s_0K_{0s_2}$ , with  $s_0$  in  $G_1$ ,  $s_2$  some element in  $G$ , and hence transforms  $a_1$  into that one of the above sets into which the coset  $K_{0s_2}$  carries  $a_1$ .

We shall prove in §24 that if an element or subgroup of an  $m$ -group  $G$  is transformed by the elements of  $G$ , the resulting set of distinct transforms constitutes a "complete set of conjugates" under  $G$ , and is transformed by the elements of  $G$  according to a transitive  $m$ -adic group of ordinary substitutions having a  $(1, N)$  isomorphism with  $G$ . We again then are concerned with the case  $\mu=2$ ; and either by applying the preceding result in conjunction with that isomorphism, or by directly extending the ordinary proof as was done above, we again obtain the complete analogue of the corresponding ordinary group theorem<sup>(50)</sup>. *A necessary and sufficient condition that a complete set of conjugate elements or subgroups under an  $m$ -group  $G$  of an element or subgroup of  $G$  is transformed under  $G$  according to an imprimitive  $m$ -adic group of ordinary substitutions is that the largest subgroup of  $G$  which transforms into itself one of these elements or subgroups is contained in a larger subgroup of  $G$ .*

When  $G$  is a transitive  $(m, \mu)$  group with  $\mu > 2$  we no longer have an analogue of  $G_1$  for  $G$  itself. We must therefore go outside of  $G$  for theorems on imprimitivity.  $G^*$  will still be transitive; and apart from the restriction  $\alpha < n$ , a set of systems of imprimitivity of either  $G$  or  $G^*$  will also be one of the other. Our description of the possible systems of imprimitivity of  $G$  therefore applies equally well to  $G^*$ , and we conclude that  $G$  will be imprimitive when and only when  $G^*$  admits a set of systems of imprimitivity for which  $\alpha < n$ . As  $G^*$  is an ordinary transitive substitution group, we easily supplement the standard result concerning its imprimitivity to obtain the following. *A transitive  $(m, \mu)$  group  $G$  is imprimitive when and only when  $G^*$  has a subgroup containing  $G_{11}^*$ , larger than  $G_{11}^*$ , but not containing  $G_0$ .  $G_{11}^*$  is of course the subgroup of  $G^*$  consisting of all of its substitutions omitting  $a_{11}$ . In proving this result we observe that as a consequence of the transitivity of  $G$  the substitutions of  $G_0$  will carry any letter into every letter in its  $\Gamma$ . If then  $G$ , and hence  $G^*$ , is imprimitive, the subgroup  $K$  of  $G^*$ , composed of all the substitutions of  $G^*$  which transform the system of imprimitivity of which  $a_{11}$  is a member into itself, satisfies the conditions of the theorem. For  $K$  is known to be a subgroup of  $G^*$  containing  $G_{11}^*$ , and larger than  $G_{11}^*$ . And as it can carry  $a_{11}$  into only  $\alpha < n$  letters of  $\Gamma_1$ , it cannot contain  $G_0$ . Conversely, if  $K$  is a subgroup of  $G^*$  satisfying the conditions of the theorem, the letters into which the substitutions of  $K$  carry  $a_{11}$  are known to form one of a set of systems of imprimitivity of  $G^*$ . As  $K$  will then contain all the substitutions of  $G^*$  which carry  $a_{11}$  into any letter that*

<sup>(50)</sup> At least as stated on page 39, *Finite Groups*.

one substitution of  $K$  carries  $a_{11}$  into, could it carry  $a_{11}$  into all the letters of  $\Gamma_1$  it would contain  $G_0$ , contrary to hypothesis. Hence, the  $\alpha$  of the resulting systems of imprimitivity of  $G$  is less than  $n$ , whence  $G$  too is imprimitive.

A criterion for the imprimitivity of  $G$  in terms of  $G_0$  would be preferable to one in terms of  $G^*$ . Our example of a primitive  $(m, 2)$  group whose associated group was imprimitive precludes such a criterion for an arbitrary  $(m, \mu)$  group. However when  $\mu = m$  we do have the following partial criterion in terms of, better than  $G_0$ , the associated constituent groups of  $G$ . *A transitive  $m$ -adic substitution group  $G$  admits systems of imprimitivity with  $\alpha > 1$  when and only when the associated constituent group  $G'_0$  is imprimitive.* In fact, systems of imprimitivity of  $G$  must be permuted as units under  $G_0$ . Hence the portions of these systems in  $\Gamma_1$  are permuted as units under  $G'_0$ . As  $\alpha > 1$  by hypothesis, and  $\alpha < n$  by definition, we thus have a set of systems of imprimitivity of  $G'_0$ . Conversely, given a set of systems of imprimitivity of  $G'_0$ , any  $s$  of  $G$  will transform  $G'_0 \rightarrow G'_0 \rightarrow \dots \rightarrow G_0^{(m-1)} \rightarrow G'_0$ , and hence will successively transform the systems of imprimitivity of  $G'_0$  into systems of imprimitivity of  $G'_0, \dots, G_0^{(m-1)}$ . The result of transforming these systems of imprimitivity of  $G_0^{(m-1)}$  by  $s$  is the same as that of transforming the given systems of imprimitivity of  $G'_0$  by  $s^{m-1}$ . As  $s^{m-1}$  is in  $G_0$ , it transforms the systems of imprimitivity of  $G'_0$  as units. Hence  $s$  transforms the totality of systems of imprimitivity of  $G'_0, G'_0, \dots, G_0^{(m-1)}$  as units. As  $G_0$  does the same, so will  $G = G_0 s$  which is therefore imprimitive with  $1 < \alpha (< n)$ .

For the exceptional case  $\alpha = 1$  we have to return to  $G^*$ . The same considerations that gave us our general criterion for the imprimitivity of a transitive  $(m, \mu)$  group yield the following result. *A transitive  $(m, \mu)$  group  $G$  admits systems of imprimitivity with  $\alpha = 1$  when and only when  $G^*$  has a subgroup containing  $G_{11}^*$ , larger than  $G_{11}^*$ , but having no other substitutions than those of  $G_{11}^*$  that carry each  $\Gamma$  into itself.* The last condition is equivalent to the crosscut of the subgroup in question and  $G_0$  being identical with  $(G_0)_{11}$ , the crosscut of  $G_{11}^*$  and  $G_0$ , and hence, for an  $m$ -adic substitution group  $G$ , with  $G_{11}^*$ .

Though all of the development of the next section, and, with certain restrictions, of the one following, can be given for  $(m, \mu)$  substitution groups, we restrict ourselves, for the sake of simplicity, to  $m$ -adic substitution groups, i.e.,  $(m, m)$  groups.

#### 19. Multiple transitivity; cyclically transitive $m$ -adic substitution groups.

Various extensions of the concept of multiple transitivity suggest themselves. According to the simplest, an  $m$ -adic substitution group  $G$  would be said to be  $r$ -fold transitive if any  $r$  letters belonging to any one  $\Gamma$  can be transformed into any  $r$  letters of the succeeding  $\Gamma$  by the substitutions of the group. It is readily proved that a necessary and sufficient condition that  $G$  be thus  $r$ -fold transitive is that the associated constituent groups  $G'_0, G'_0, \dots, G_0^{(m-1)}$  (or any one of them) be  $r$ -fold transitive in the ordinary sense. Since the order of  $G'_0$  is a divisor of the order of  $G_0$ , and hence of  $G$ , it follows from the corre-



sponding ordinary group result that the order of an  $r$ -fold transitive  $m$ -adic substitution group of degree  $n$  is a multiple of  $n(n-1) \cdots (n-r+1)$ .

Of special interest in polyadic theory is the type of multiple transitivity we term cyclic transitivity. An  $m$ -adic substitution group  $G$  will be said to be *cyclically transitive* if, given any two selections from the classes of letters  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$ , some substitution of  $G$  will carry the letters of one selection into the letters of the other. Actually then, if one selection is  $a_{1j_1}, a_{2j_2}, \dots, a_{(m-1)j_{m-1}}$ , the other  $a_{1k_1}, a_{2k_2}, \dots, a_{(m-1)k_{m-1}}$ , the substitution will carry  $a_{1j_1} \rightarrow a_{2k_2}, a_{2j_2} \rightarrow a_{3k_3}, \dots, a_{(m-1)j_{m-1}} \rightarrow a_{1k_1}$ . Every cyclically transitive  $m$ -adic substitution group is then transitive, and, indeed, for  $m=2$  cyclic transitivity reduces to transitivity. The symmetric and alternating  $m$ -adic groups of degree  $n$ , previously observed to be transitive—in the latter cases at least for  $n > 2$ —are now seen to be cyclically transitive.

The  $m-1$   $\Gamma$ 's, of  $n$  letters each, give rise to  $n^{m-1}$  selections which we shall call *cycles*. Any  $m$ -adic substitution  $s$  on the  $\Gamma$ 's will merely permute these cycles, and hence will determine an ordinary substitution on these  $n^{m-1}$  cycles as new "permutants." Since this relationship is preserved under multiplication, the members of an  $m$ -adic substitution group  $G$  of degree  $n$  will thus give rise to substitutions on the cycles forming an  $m$ -adic group  $G'$  of ordinary substitutions, of degree  $n^{m-1}$ , isomorphic with  $G$ . Clearly, different  $m$ -adic substitutions yield different substitutions of the cycles. Hence  $G'$  is indeed simply isomorphic with  $G$ . In particular, then,  $G'$  and  $G$  are of the same order. Finally, if  $G$  is cyclically transitive,  $G'$  will be transitive, and, indeed, conversely. Since the order of an  $m$ -adic transitive group of ordinary substitutions is a multiple of its degree, we have, as our first result, *the order of a cyclically transitive  $m$ -adic substitution group of degree  $n$  is a multiple of  $n^{m-1}$ .*

We have seen that transitive  $m$ -adic groups of ordinary substitutions have the complete analogue of the  $G_1$  of ordinary transitive groups. Actually, all of the corresponding theory goes over. In our simple isomorphism between  $G$  and  $G'$ , the subgroup of  $G'$  consisting of all of its substitutions "omitting" a given cycle  $C$  will correspond to the subgroup  $G_C$  of  $G$  consisting of all of its substitutions which carry  $C$  into itself. We shall call  $G_C$  the *cycle subgroup* of  $G$ , corresponding to  $C$ . Actually, if  $C$  is the selection  $a_{1j_1}, a_{2j_2}, \dots, a_{(m-1)j_{m-1}}$ , it will be transformed into itself according to the cyclic substitution  $(a_{1j_1} a_{2j_2} \cdots a_{(m-1)j_{m-1}})$  by all the substitutions of  $G_C$ . Hence, if the  $m$ -adic substitutions of  $G$  be written as ordinary substitutions on  $(m-1)n$  letters in cycle form,  $G_C$  will consist of those substitutions of  $G$  which have this cyclic substitution as component. It will be convenient to speak of these substitutions as having the cycle  $C$ . Clearly, then, *an  $m$ -adic substitution of degree  $n$  cannot have more than  $n$  cycles.*

Each of the  $n^{m-1}$  cycles yields thus a corresponding cycle subgroup of the cyclically transitive  $G$ . The simple isomorphism between  $G$  and  $G'$  then immediately transforms the corresponding properties of  $G'$  to yield, among



others, the following results on  $G$ . The order of each cycle subgroup of  $G$  is equal to the order of  $G$  divided by  $n^{m-1}$ . The cycle subgroups of  $G$  are conjugate, forming a complete set of conjugates under the substitutions of  $G$ . If all the substitutions of a given cycle subgroup have exactly  $\alpha$  cycles in common, and they have one cycle in common by definition, then the  $n^{m-1}$  cycles can be separated into mutually exclusive sets of  $\alpha$  cycles each such that different cycles yield the same cycle subgroup when and only when they belong to the same set.

There are thus  $n^{m-1}/\alpha$  distinct cycle subgroups of  $G$ . The only information that  $G'$  yields concerning  $\alpha$  is that it is a divisor of  $n^{m-1}$ . However, our observation that an  $m$ -adic substitution of degree  $n$  cannot have more than  $n$  cycles shows that  $\alpha \leq n$ . Hence, a cyclically transitive  $m$ -adic substitution group of degree  $n$  has a number  $N$  of cycle subgroups with  $N \geq n^{m-2}$  and a divisor of  $n^{m-1}$ . Whether  $N$  is actually a multiple of  $n^{m-2}$ , i.e.,  $\alpha$  a divisor of  $n$ , is another of our unsolved problems.

**20. Class of an  $m$ -adic substitution group.** The class of an ordinary substitution group is the smallest number of letters appearing in any of its substitutions, other than the identity, when those substitutions are written in cycle form. Since the substitutions of an  $m$ -adic substitution group  $G$  never carry a letter into itself when  $m > 2$ , we are led to define the class of  $G$  as the class of its associated group  $G_0$ . This also is the class of its containing group  $G^{(*)}$ .

With this definition most of the elementary theory of class goes over to  $m$ -adic substitution groups. We have almost immediately that the  $m$ -adic symmetric group of degree  $n$ ,  $n > 1$ , is a primitive group of class 2, while the  $m$ -adic alternating groups of degree  $n$ ,  $n > 2$ , are primitive groups of class 3. That these  $m$ -adic groups are primitive follows from the fact that their constituent associated groups are either the symmetric, or alternating, ordinary groups of degree  $n$ , and hence primitive, so that the  $m$ -adic groups do not admit systems of imprimitivity with  $\alpha > 1$ ; and, being cyclically transitive, they cannot admit systems of imprimitivity with  $\alpha = 1$ . As for their class, they are clearly at most of the class indicated. And could an alternating group actually be of class 2, the  $\epsilon$ -subgroup of its corresponding  $\delta$ -subgroup would have an  $\epsilon$ -sequence with one  $-1$ ; but then the  $\delta$ -subgroup would be the complete  $\delta$ -group, and hence the given group not an alternating group, but the symmetric group.

We now prove that, as in the standard theory, the converses of these results also hold. First then let  $G$  be a primitive  $m$ -adic substitution group of degree  $n$  and of class 2. On the one hand, its associated constituent group  $G'$  will be primitive; on the other hand, its associated group  $G_0$  will have some substitution whose component in each  $G_0^{(i)}$  but one is the identity, and in

(<sup>60</sup>) Not necessarily so, however, for  $(m, \mu)$ -groups with  $\mu < m$ .

that one a transposition. By the invariance of  $G_0$  under  $G$ , we see that  $G_0$  has a substitution  $t_0$  of the form  $t'_0 \cdot 1 \cdots 1$ , with  $t'_0$  in  $G'_0$  and a transposition. Now let  $\bar{G}'_0$  be that subgroup<sup>(a1)</sup> of  $G'_0$  composed of all the substitutions  $t'$  of  $G'_0$  for which  $t' \cdot 1 \cdots 1$  is in  $G_0$ . The subgroup  $\bar{G}'_0$  is clearly an invariant subgroup of  $G_0$ , and hence of  $G'_0$ , and it has the transposition  $t'_0$ . Now the standard proof of the fact that a primitive (ordinary) group of class 2 is the corresponding symmetric group also yields the following more general statement. An invariant subgroup<sup>(a2)</sup> of class 2 of a primitive group is the corresponding symmetric group. Hence  $\bar{G}'_0$  is the symmetric group of degree  $n$ .  $G_0$  therefore has among its elements every substitution of the form  $t' \cdot 1 \cdots 1$ . Since  $G_0$  is invariant under  $G$ , it also has every substitution of the form  $1 \cdots t^{(i)} \cdots 1$ , for each  $i$ , and hence every substitution of the form  $t' t'' \cdots t^{(m-1)}$ .  $G_0$  is therefore the associated group of the  $m$ -adic symmetric group of degree  $n$ , and hence  $G$  the symmetric group itself. Hence, *every primitive  $m$ -adic substitution group of class 2 and degree  $n$  is the corresponding symmetric group, and conversely for  $n > 1$ .*

If  $G$  is a primitive  $m$ -adic substitution group of degree  $n$  and class 3, we have as before that  $G'_0$  is primitive, while  $G_0$  has a substitution of the form  $t'_0 \cdot 1 \cdots 1$  with  $t'_0$  of the form  $abc$ , the last since the substitution of class 3 in  $G_0$  must consist of a single cycle of three letters which, in turn, must then belong to a single  $\Gamma$ . Defining  $\bar{G}'_0$  as before, we see that  $\bar{G}'_0$  is of class 3, and hence, by the corresponding extension of the standard result, is the alternating group of degree  $n$ . We therefore conclude that  $G_0$  has, perhaps among others, every substitution of the form  $t' t'' \cdots t^{(m-1)}$  with the  $t^{(i)}$ 's positive substitutions.  $G_0$  therefore has every possible substitution corresponding to the  $\epsilon$ -sequence  $(+1, +1, \cdots, +1)$ , and hence every possible substitution for each of the  $\epsilon$ -sequences of its substitutions. It is therefore the associated group of an alternating group, i.e.,  $G$  is an alternating group. Hence, *every primitive  $m$ -adic substitution group of degree  $n$  and of class 3 is an alternating group of degree  $n$ , and conversely for  $n > 2$ .*

Actually two cases arise as far as  $G'_0$  is concerned. When the above found substitutions of  $G_0$  are its only substitutions,  $(+1, +1, \cdots, +1)$  is its only  $\epsilon$ -sequence,  $G$  is an alternating group whose  $\delta$ -subgroup is of the first order, while  $G'_0$  is identical with  $\bar{G}'_0$ , and hence is itself the alternating group. Otherwise,  $G'_0$  will be larger than  $\bar{G}'_0$ , while containing it, and hence will be the symmetric group, while  $G$  will be an alternating group whose  $\delta$ -subgroup is of order greater than one. Note also that in both of the above results the hypothesis of the primitivity of  $G$  was used only in deducing the primitivity of  $G'_0$ . We therefore conclude that there does not exist an  $m$ -adic substitution group  $G$  of class 2 or 3 for which  $G$  is imprimitive,  $G'_0$  primitive.

(a1) If not  $G'_0$  itself.

(a2) Actually improper, therefore.

Let now  $G$  be a primitive  $m$ -adic group of degree  $n$  and of class  $p$ ,  $p$  a prime greater than 3. As before, the substitution of class  $p$  in  $G_0$  must consist of a single cycle of letters which therefore belong to a single  $\Gamma$ . Hence  $n \geq p$ . Furthermore,  $\bar{G}_0'$  will have a substitution of class  $p$  for element, and hence be of class  $p$ . Finally  $G_0'$  is primitive, with  $\bar{G}_0'$  as invariant subgroup. With the corresponding ordinary proof generalized as in the preceding cases, we then find that  $\bar{G}_0'$  is  $(n-p+1)$ -fold transitive. The remainder of the standard proof is then directly applicable to  $\bar{G}_0'$  and shows that  $n$  cannot be greater than  $p+2$ . Hence, if a primitive  $m$ -adic substitution group is of class  $p$ ,  $p$  being a prime number greater than 3, its degree can only be  $p$ ,  $p+1$  or  $p+2$ . Note that actually  $\bar{G}_0'$  is then itself primitive—immediately so for  $n=p$ , and as a consequence of its being more than simply transitive for  $n=p+1$  or  $p+2$ . Hence, in each of these cases,  $\bar{G}_0'$  is the unique primitive ordinary group of class  $p$  and degree  $n$ .

We consider in detail only the case  $n=p$ .  $\bar{G}_0'$  is then the group of order  $p$ , as is also each  $\bar{G}_0^{(i)}$ , defined in analogous fashion. Each  $\bar{G}_0^{(i)}$  is therefore a cyclic group, and is, in fact, generated by a single cycle of the  $p$  letters of  $\Gamma_i$ . By relettering the members of the  $\Gamma$ 's we may therefore assume that  $\bar{G}_0^{(i)}$  is generated by the substitution  $t_0^{(i)} = (a_{i1}a_{i2} \cdots a_{ip})$ . Now any substitution  $t$  of  $G_0$  will transform each  $\bar{G}_0^{(i)}$  into itself, and hence will transform  $t_0^{(i)}$ , the generator of  $\bar{G}_0^{(i)}$ , into some power  $\nu^{(i)}$  of itself, with  $\nu^{(i)} = 1, 2, \dots, p-1$ . Hence, with each  $t$  in  $G_0$  we can thus associate a  $\nu$ -sequence  $(\nu', \nu'', \dots, \nu^{(m-1)})$ . Likewise, if  $s$  is any substitution in  $G$ ,  $s$  will transform each  $\bar{G}_0^{(i)}$  into  $\bar{G}_0^{(i+1)}$ . It will therefore transform each  $t_0^{(i)}$  into some power  $\mu^{(i)}$  of  $t_0^{(i+1)}$ ,  $\mu^{(i)} = 1, 2, \dots, p-1$ . Hence, with each  $s$  in  $G$  we can thus associate a  $\mu$ -sequence  $(\mu', \mu'', \dots, \mu^{(m-1)})$ . Since  $G_0$  has the substitution  $1 \cdots t^{(i)} \cdots 1$  whenever  $\bar{G}_0^{(i)}$  has the substitution  $t^{(i)}$ , we see that  $G_0$  has the invariant subgroup

$$\bar{G}_0 = \bar{G}_0' \bar{G}_0'' \cdots \bar{G}_0^{(m-1)},$$

the direct product of the  $\bar{G}_0^{(i)}$ 's, when it is not  $\bar{G}_0$  itself. Now  $\bar{G}_0^{(i)}$  consists of all the substitutions on the letters of  $\Gamma_i$  that transform  $t_0^{(i)}$  into itself. It follows that  $\bar{G}_0$  consists of all the  $(m-1)$ -ads, consequently  $p^{m-1}$  in number, which transform each  $t_0^{(i)}$  into itself.  $G_0$ , therefore, has among its elements each of the  $p^{m-1}$   $(m-1)$ -ads with which we can associate the  $\nu$ -sequence  $(1, 1, \dots, 1)$ . By expanding  $G_0$  in cosets as regards  $\bar{G}_0$ , we then easily verify that  $G_0$  likewise has each of the  $(m-1)$ -ads with which we can associate the  $\nu$ -sequence of any one of its members, there being exactly  $p^{m-1}$   $(m-1)$ -ads for each  $\nu$ -sequence. Likewise, by expanding  $G$  in cosets as regards  $\bar{G}_0$ , which is invariant under  $G$ , we find that  $G$  has every  $m$ -adic substitution on the  $\Gamma$ 's with which we can associate the  $\mu$ -sequence of any one of its members, there being exactly  $p^{m-1}$  such substitutions for each  $\mu$ -sequence.

$G$  is therefore determined by the set of  $\mu$ -sequences of its members. Actually, if  $s_i$  in  $G$  has the  $\mu$ -sequence  $(\mu_i', \mu_i'', \dots, \mu_i^{(m-1)})$ , then  $s = s_1 s_2 \cdots s_m$ ,



"alternating power group," the  $m$ -adic substitution group of degree  $p$  consisting of all the  $m$ -adic substitutions on the  $\Gamma$ 's with  $\mu$ -sequences in the  $\mu$ -subgroup under consideration. Each of our  $G$ 's is therefore an alternating power group<sup>(63)</sup>. To complete our investigation within its present scope we need merely find which of the alternating power groups are primitive groups of class  $p$ . Actually they are all primitive. For their  $\bar{G}'_0$  is the primitive  $\bar{P}'_0$ , so that their  $G'_0$  is primitive. And their  $\bar{G}_0$  is always  $\bar{P}_0$ , which can carry any selection of letters chosen from the  $\Gamma$ 's into any other selection, so that in fact, they are cyclically transitive. As for their class, it is immediately seen to be at most  $p$ . Now actually a substitution on the letters of  $\Gamma$ , carrying  $t_0^{(0)} = (a_{11}a_{12} \cdots a_{1p})$  into a power of itself other than the first must be of class  $p-1$ . It follows that an alternating power group of degree  $p$  is of class less than  $p$ , in fact  $p-1$ , when and only when the associated group of its  $\mu$ -subgroup has a  $\nu$ -sequence with one and only one number not unity. This is easily transformed into a condition on the  $\mu$ -subgroup itself to yield the following result. *The primitive groups of class  $p$  and degree  $p$ ,  $p$  being a prime greater than 3, are the alternating power groups of degree  $p$  whose  $\mu$ -subgroups do not have a pair of  $\mu$ -sequences differing in one and only one component<sup>(64)</sup>.*

#### B. FINITE ABSTRACT POLYADIC GROUPS

21. **Cyclic polyadic groups; ordinary theory<sup>(65)</sup>.** Given the  $m$ -adic operation  $c$ , we define the  $m$ -adic powers of an element  $s$  under  $c$  inductively as follows.  $s$  itself will be rewritten  $s^{[0]}$ ; and having  $s^{[n]}$ , we define  $s^{[n+1]}$  as  $c(s \cdots ss^{[n]})$ . If then  $s^{[n]}$  be written out in full,  $n$  is the number of  $c$ 's occurring in the resulting extended operation, the number of  $s$ 's being  $n(m-1)+1$ . By the associative law it follows that any extended operation involving  $n$   $c$ 's and but the single element  $s$  repeated can be rewritten in the form  $s^{[n]}$ . We thus easily obtain the following  $m$ -adic power laws:

$$c(s^{[n_1]}s^{[n_2]} \cdots s^{[n_m]}) = s^{[n_1+n_2+\cdots+n_m+1]}, \quad (s^{[n_1]})^{[n_2]} = s^{[(m-1)n_1n_2+n_1+n_2]}.$$

Note that for  $m=2$  our  $n$ th power is the ordinary  $(n+1)$ -st power<sup>(66)</sup>.

<sup>(63)</sup> Unless it were  $P$  itself. But  $P$  is readily seen to be of class  $p-1$ .

<sup>(64)</sup> The actual problem of determining the subgroups of the complete  $\mu$ -group remains unsolved. Gill has pointed out to the writer that while the problem of determining the associated groups of these  $\mu$ -subgroups can superficially be expressed as a problem in V.A.G.'s, actually the theory is now inapplicable, since the coefficients of the polynomials no longer form a field.

<sup>(65)</sup> For the special case  $m=3$ , the results of the present section reduce to those given by Lehmer. Likewise those of the next section involving mere reducibility, now of necessity to a 2-group.

<sup>(66)</sup> By contrast, Dörnte writes  $a^z$  in usual notation with, however,  $z$  subject to the restriction  $z \equiv 1 \pmod{m-1}$ . While our laws of powers are, as a result, more complicated than Dörnte's, we find great comfort in the fact that our  $s^{[n]}$  is an " $m$ -adic element" for every positive integral, or zero,  $n$ . Our lack of negative  $m$ -adic powers could easily be supplied.



If  $s$  is an element of an  $m$ -adic group  $K$ , each of its  $m$ -adic powers will represent elements of  $K$ . With  $K$  a finite group we therefore must have for some  $n_0$  and  $n_0 + n$ ,  $n > 0$ ,  $s^{[n_0]} = s^{[n_0+n]}$ . Since  $s^{[n_0+n]}$  can be rewritten  $c(s^{[n_0]}s \cdots ss^{[n-1]})$ , it follows that  $\{s, \cdots, s, s^{[n-1]}\}$  is an identity, whence we have

$$s^{[n]} = s.$$

The smallest positive integral value of  $n$  for which this equation holds will be called the ( $m$ -adic) *order* of  $s$ . If then  $s$  is of order  $g$ , the sequence of its  $m$ -adic powers  $s^{[0]}, s^{[1]}, s^{[2]}, \cdots$  starts with  $g$  distinct elements which are then repeated in order. It follows on the one hand that  $s^{[n]} = s$  when and only when  $n$  is a multiple of  $g$ ; and, more generally, that  $s^{[n_1]} = s^{[n_2]}$  when and only when  $n_1 - n_2$  is a multiple of  $g$ . On the other hand, since but a finite number of elements are involved, our first law of  $m$ -adic powers shows that the  $g$  distinct elements constitute an abelian  $m$ -adic group  $G$  of order  $g$  which may then be called the *cyclic  $m$ -adic group generated by  $s$* . The order of  $s$  is therefore equal to the order of the cyclic group it generates. Again by the first law of  $m$ -adic powers it is immediately seen that two cyclic  $m$ -groups of the same order are simply isomorphic. Furthermore, the same law shows that apart from an assumed  $m$ -group  $K$ ,  $g$  distinct elements  $s_0, s_1, \cdots, s_{g-1}$ , subject to the  $m$ -adic operation obtained by writing  $s_n = s^{[n]}$ , with  $s^{[0]} = s$ , constitute an  $m$ -group which is then the cyclic  $m$ -group of order  $g$  generated by  $s = s_0$ . Hence, as in ordinary group theory, we may say there is one and only one cyclic  $m$ -group whose order is an arbitrary natural number<sup>(67)</sup>.

Let then  $G$  be the cyclic  $m$ -group of order  $g$ ,  $s$  a generator of  $G$ . We first ask for the order of any power  $s^{[n]}$  of  $s$ . This will be the least value of  $N$  for which  $(s^{[n]})^{[N]} = s^{[n]}$ , hence the least value of  $N$  for which  $(m-1)nN + N + n - n = [(m-1)n+1]N$  is a multiple of  $g$ . It follows that the order of  $s^{[n]}$  is equal to the order of  $s$  divided by the highest common factor of  $(m-1)n+1$  and the order of  $s$ . In particular, the order of  $s^{[n]}$  will be the same as the order of  $s$  when and only when  $(m-1)n+1$  is prime to the order of  $s$ . Hence an element  $s$  is generated by those and only those of its  $m$ -adic powers  $s^{[n]}$  for which  $(m-1)n+1$  is prime to the order of  $s$ .

We can now determine what orders the elements of  $G$  can have.  $\gamma$  will be the order of an element of  $G$  if  $\gamma = g/d$ ,  $d = \text{H.C.F.} [(m-1)n+1, g]$  for some  $n$ . It is necessary then that  $d$  be a divisor of  $g$ , and prime to  $m-1$ . We now show that this is also sufficient. We have to find, then, an  $n$  and  $k$  such that  $(m-1)n+1 = kd$ ,  $g = \gamma d$  with  $k$  relatively prime to  $\gamma$ . Since  $m-1$  is prime to  $d$  by hypothesis, for some  $n = n_0$ ,  $k = k_0$ , we will have  $(m-1)n_0+1 = k_0d$ .

<sup>(67)</sup> The following discussion tacitly assumes that a symbol representing the order of an element or group is restricted to positive integral values, one representing an  $m$ -adic power to non-negative integral values. On the other hand, symbols entering into a diophantine equation may at first be allowed to assume arbitrary integral values which are then restricted in the above manner as the need arises.



The general solution of  $(m-1)n+1=kd$  is then given by  $n=n_0+\lambda d$ ,  $k=k_0+\lambda(m-1)$  with arbitrary  $\lambda$ . Now the particular solution shows  $k_0$  to be prime to  $m-1$ . Hence the arithmetic progression  $k_0+\lambda(m-1)$  has, indeed, an infinite number of primes, and hence certainly a number prime to  $\gamma$  as was to be proved. We thus have the following result. *A cyclic  $m$ -group of order  $g$  has at least one element of every order  $\gamma$  such that  $\gamma$  is a divisor of  $g$ , and  $g/\gamma$  is prime to  $m-1$ , and no element of any other orders. In particular, a cyclic  $m$ -group of order  $g$  has a first order element when and only when  $g$  is prime to  $m-1$ .*

We can now generalize the ordinary cyclic group argument to prove the following. *A cyclic  $m$ -group of order  $g$  has one and only one subgroup whose order is any given divisor  $\gamma$  of  $g$  such that  $g/\gamma$  is prime to  $m-1$ , and no others.* The one subgroup is immediately yielded by the cyclic subgroup generated by an element of order  $\gamma$ , whose existence is insured by the preceding result. For the converse, consider any subgroup of the given cyclic group, and let its order be  $\gamma$ . By Lagrange's theorem extended,  $\gamma$  is a divisor of  $g$ . By the same theorem, each element  $s$  of the subgroup has an order which is a divisor of  $\gamma$ , and hence must satisfy the equation  $s^{[\gamma]}=s$ . Now consider all the elements of the given cyclic group, generated, say, by  $s_0$ , that satisfy this equation. If  $s=s_0^{[n]}$ , we have, as in a preceding argument, that  $\gamma[(m-1)n+1]=kg$  for some  $k$ , and hence, with  $g/\gamma=d$ , that  $(m-1)n+1=kd$ —and conversely. We first see that  $d$  is prime to  $m-1$ , and hence that  $\gamma$  is the order of a cyclic subgroup of the given group. Furthermore, since  $\gamma d=g$ , our general solution  $n=n_0+\lambda d$ ,  $k=k_0+\lambda(m-1)$ , of the equation  $(m-1)n+1=kd$  shows that exactly  $\gamma$  such  $n$ 's are to be found with values in the set  $0, 1, 2, \dots, g-1$ . Hence the elements  $s$  satisfying the equation  $s^{[\gamma]}=s$  are exactly  $\gamma$  in number, and consequently must be the  $\gamma$  elements of the above cyclic subgroup of order  $\gamma$ . Our assumed subgroup of order  $\gamma$  must therefore be that cyclic subgroup, whence our result.

From this proof flow a number of corollaries. We have immediately that *every subgroup of a cyclic polyadic group is cyclic*. Furthermore, our proof shows that an element of given order of a cyclic group is contained in those and only those subgroups of the cyclic group whose orders are multiples of the order of the element. It follows, on the one hand, that the necessary and sufficient condition that one element of a cyclic group generate a second is that the order of the first be a multiple of the order of the second. On the other hand, we see that two subgroups of a cyclic polyadic group intersect in the subgroup whose order is the highest common factor of the orders of the given subgroups, and generate the subgroup whose order is the least common multiple of those orders.

Apart from the possible orders of elements and subgroups of a cyclic polyadic group the above results are the same as for ordinary cyclic groups. Our condition on those possible orders  $\gamma$  can be transformed into the following more usable form. *Let  $g_0$  be the largest divisor of  $g$  prime to  $m-1$ , and let*

$\gamma_0 = g/g_0$ . Then the cyclic  $m$ -group of order  $g$  has at least one element, and exactly one subgroup, of those and only those orders  $\gamma$  for which  $\gamma = \delta\gamma_0$ ,  $\delta$  a divisor of  $g_0$ . In fact, if  $\gamma$  is a divisor of  $g$  with  $g/\gamma$  prime to  $m-1$  as per our original condition,  $g/\gamma$ , being a divisor of  $g$  prime to  $m-1$ , must be a divisor of  $g_0$ . Hence  $g_0 = \delta(g/\gamma)$  with  $\delta$  a divisor of  $g_0$ , whence  $\gamma = \delta\gamma_0$ . Conversely, if  $\gamma = \delta\gamma_0$  with  $\delta$  a divisor of  $g_0$ ,  $g/\gamma = g_0/\delta$ , so that  $\gamma$  is a divisor of  $g$  with  $g/\gamma$  prime to  $m-1$ .

We thus see that  $\gamma_0$  is the least order of a subgroup of our cyclic group, with all of the subgroups of the cyclic group containing the unique subgroup of order  $\gamma_0$ . At one extreme, when  $\gamma_0 = 1$ , which is equivalent to  $g$  prime to  $m-1$ , the cyclic group has a subgroup of first order. This corresponds to the element of first order previously noted, which is now seen to be unique. Every subgroup then contains this first order element, and their orders are the same as the orders of the subgroups of an ordinary cyclic group of order  $g$ . At the other extreme  $\gamma_0 = g$ , which is equivalent to every distinct prime factor of  $g$  being a factor of  $m-1$ . The cyclic group then has no (proper) subgroup, each of its elements being of order  $g$ , and thus generating the entire cyclic group<sup>(68)</sup>. In particular, if  $g$  is a prime  $p$ , the corresponding cyclic group is always of one of these two special types. Note that unlike an ordinary group, a polyadic group whose order is a prime  $p$  need not be cyclic. By the extended Lagrange theorem its elements must be of order 1 or  $p$ . If it has an element of order  $p$ , it must be the cyclic group of order  $p$ . However all of its elements may be of order one, in which case it is noncyclic.

In the general case the orders of the subgroups are multiples of  $\gamma_0$ , the multipliers being the orders of the subgroups of an ordinary cyclic group of order  $g_0$ . Hence, if  $g = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_p^{\alpha_p} q_1^{\beta_1} q_2^{\beta_2} \cdots q_\sigma^{\beta_\sigma}$ , with  $p_1, p_2, \dots, p_p$  distinct primes not factors of  $m-1$ ,  $q_1, q_2, \dots, q_\sigma$  factors of  $m-1$ , then the number of subgroups of the cyclic  $m$ -group of order  $g$  is  $(\alpha_1+1)(\alpha_2+1) \cdots (\alpha_p+1) - 1$ .

We can now also find an expression for the number of elements of given order in a cyclic  $m$ -group. With that order one for which there is at least one element, the total number of elements of that order will be the same as the number of generators of a cyclic  $m$ -group of that order. We proceed therefore to find the number of generators of a cyclic  $m$ -group of order  $g$  generated, say, by  $s_0$ . We first show that if  $m-1$  is prime to  $g$  the number of generators is  $\phi(g)$  as for ordinary cyclic groups. In fact, if we recall our formula for the order of  $s_0^{[n]}$  we see that the number of generators in question is the number of numbers  $(m-1)n+1$ ,  $n=0, 1, \dots, g-1$ , prime to  $g$ . But with  $m-1$  prime to  $g$  this is the same as the number of numbers  $0, 1, \dots, g-1$  prime to  $g$ , that is,  $\phi(g)$ . Now, in the general case, expand the given cyclic  $m$ -group of order  $g$  in cosets as regards its subgroup of order  $\gamma_0$ . The resulting quotient group is then an  $m$ -group of order  $g_0$  prime to  $m-1$ . Now let  $s$  be any element of the given group,  $\sigma$  the corresponding element of the quotient group. Then  $s$

(68) Whence our correction of a statement of Miller.

is a generator of the given group when and only when  $\sigma$  is a generator of the quotient group. That  $\sigma$  generates the quotient group if  $s$  generates the given group is immediate. As for the converse,  $s$  will then generate a group having a complete set of multipliers for our coset expansion. But it must also generate all the elements of the subgroup of order  $\gamma_0$ , and hence all the elements of the group. It follows, on the one hand, that the quotient group is itself cyclic, and hence has  $\phi(g_0)$  generators, and hence, finally, that the number of generators of a cyclic  $m$ -group of order  $g$  is  $\gamma_0 \phi(g_0)$ .

Among the few extensions of topics of the ordinary theory of cyclic groups omitted in the above development is that of the  $k$ th powers of elements of a cyclic group. We state the result for  $m$ -groups without further proof. *The distinct  $k$ th powers of a cyclic  $m$ -group of order  $g$  constitute a subgroup of order  $g/h$  where  $h$  is the highest common factor of  $g$  and  $(m-1)k+1$ ; furthermore, each element of this subgroup is the  $k$ th power of exactly  $h$  elements of the given group.*

**22. Cyclic polyadic groups; polyadic theory.** We have observed that a cyclic  $m$ -group of order  $g$  has a first order element when and only when  $m-1$  is prime to  $g$ . As this element, when it exists, is invariant under the group, it follows that a cyclic  $m$ -group of order  $g$  is reducible to a 2-group when and only when  $g$  is prime to  $m-1$ . We turn now to the general discussion of reducibility for cyclic polyadic groups. Our first result is immediate. *Every group to which a cyclic group is reducible is cyclic.* For if  $s$  is a generator of the given cyclic group,  $c$  its operation,  $c'$  the operation of the reduced group, every element of the given group is given by an extended  $c$  operation involving  $s$ 's only, hence also by an extended  $c'$  operation involving  $s$ 's only.  $s$  is therefore a generator of the reduced group, which is thus cyclic.

In applying our general criterion of reducibility to cyclic groups, questions of commutativity are automatically disposed of, since every cyclic group is abelian. A cyclic  $m$ -group will then be reducible to a  $\mu$ -group,  $m = k(\mu-1) + 1$ , if for some  $(\mu-1)$ -ad, which may be written  $\{s^{[n_1]}, s^{[n_2]}, \dots, s^{[n_{\mu-1}]}\}$ ,  $s$  being a generator of the cyclic group, the  $(m-1)$ -ad  $\{s^{[n_1]}, s^{[n_2]}, \dots, s^{[n_{\mu-1}]}, \dots, s^{[n_1]}, s^{[n_2]}, \dots, s^{[n_{\mu-1}]}\}$  is an identity of the cyclic group. Hence also if  $s^{[k(n_1+n_2+\dots+n_{\mu-1})+1]} = s$ , i.e., if  $kn+1$  is a multiple of  $g$ , where  $g$  is the order of the group,  $n = n_1 + n_2 + \dots + n_{\mu-1}$ . It follows first that the cyclic  $m$ -group is reducible to a  $\mu$ -group when and only when  $k = (m-1)/(\mu-1)$  is prime to  $g$ . Furthermore, with  $k$  prime to  $g$ , if  $kn'+1$  and  $kn''+1$  are both multiples of  $g$ , then  $k(n'-n'')$ , and hence  $n'-n''$ , must be a multiple of  $g$ . It follows from our first law of  $m$ -adic powers that the  $(\mu-1)$ -ads corresponding to  $n'$  and  $n''$  are equivalent. Recalling our general theory of reducibility, we thus see that a cyclic  $m$ -group is reducible to but one  $\mu$ -group for each admissible  $\mu$ .

The least value of  $\mu$  for which  $(m-1)/(\mu-1)$  is prime to  $g$  corresponds to a  $k$  which is the largest divisor of  $m-1$  prime to  $g$ . We thus obtain the following result. *The real dimension of a cyclic  $m$ -group of order  $g$  is equal to  $(m-1)/k_0+1$  where  $k_0$  is the largest divisor of  $m-1$  prime to  $g$ . In particular*

a cyclic  $m$ -group of order  $g$  is irreducible when and only when each prime factor of  $m-1$  is also a prime factor of  $g$ . Our previous uniqueness result easily enables us to complete the picture as far as mere reducibility is concerned. We thus see that the real dimension of a cyclic  $m$ -group of order  $g$  is its only irreducible dimension; and the groups to which the cyclic  $m$ -group is reducible, all cyclic, consist of a single group of dimension equal to the real dimension of the given group, and those of its extensions whose dimensions are of the form  $k(\mu_0-1)+1$ , where  $\mu_0$  is the real dimension in question,  $k$  any proper divisor of  $k_0 = (m-1)/(\mu_0-1)$ .

Since the subgroups of a cyclic group are themselves cyclic, we can find their real dimensions by applying the above formula.  $\gamma$  will be the order of a subgroup of a cyclic  $m$ -group of order  $g$  if it is a divisor of  $g$  with  $g/\gamma$  prime to  $m-1$ . Writing  $g = d\gamma$ , with  $d$  prime to  $m-1$ , we see that the largest divisor of  $m-1$  prime to  $\gamma$  is also the largest divisor of  $m-1$  prime to  $g$ . Hence, all the subgroups of a cyclic polyadic group have the same real dimension, namely the real dimension of the group itself. It follows that a subgroup of a cyclic  $m$ -group is reducible to a  $\mu$ -group when and only when the given group is reducible to a  $\mu$ -group. In particular, all the subgroups of an irreducible cyclic group are irreducible. We now readily verify that the following simple situation holds. If a cyclic  $m$ -group be reduced to a  $\mu$ -group, the subgroups of the  $m$ -group are thereby reduced to the subgroups of the  $\mu$ -group<sup>(69)</sup>. In fact, half of this situation obtains for arbitrary polyadic groups. For from the very definition of reducibility, if a polyadic group  $G$  is reduced to a polyadic group  $G'$ , the subgroups of  $G'$  are also subgroups of  $G$ , or more exactly, reductions of subgroups of  $G$ . Moreover, if  $G$  is abelian, every reduction of a subgroup of  $G$  can be effected by thus reducing  $G$  to some  $G'$ . For the satisfaction of our general criterion of reducibility by the subgroup then holds equally well for  $G$ , and the same operation that serves to reduce the subgroup is shown by the proof of that criterion to reduce  $G$  as well. If now  $G$  is cyclic and reducible to a  $\mu$ -group, every subgroup of  $G$  is reducible to a  $\mu$ -group; and since  $G$  can be reduced to but a single  $\mu$ -group  $G'$ , that reduction must reduce all the subgroups of  $G$  to subgroups of  $G'$ , and hence by the first part of the proof to the subgroups of  $G'$ . Our proof incidentally shows that by varying  $\mu$  every possible reduction of a subgroup of  $G$  will thus be obtainable. We furthermore have the following corollary which, indeed, can easily be proved directly, and itself used to give a different turn to our proofs. The polyadic orders of the elements of a cyclic polyadic group remain unchanged under every reduction of the group.

While cyclic groups form a closed set with respect to the two operations "subgroup of" and "reduction of," they do not form a closed set under the operation of extension, of which reduction is the inverse, and hence under the more general operation of derivation. We proceed to prove that a cyclic

<sup>(69)</sup> We prove this result independently of the discussion of complexes which concluded §5, since that discussion was extremely sketchy.

*m*-group of order *g* remains cyclic when extended to a  $\mu$ -group,  $\mu = k(m-1) + 1$ , when and only when *k* is prime to *g*. Let the polyadic *n*th powers of an element *s* in the two groups be written more explicitly  $s^{[n]_m}$  and  $s^{[n]_\mu}$ . By counting *c*'s we then have immediately

$$s^{[n]_\mu} = s^{[kn]_m}.$$

Let *s* be a generator of the extended group, assuming that group to be cyclic. The elements of that group, and hence of the given group, will then be given by the  $\mu$ -adic powers of *s*, and hence by those *m*-adic powers of *s* of the form  $s^{[kn]_m}$ . For each *N*, therefore, there must be an *n* such that  $s^{[N]_\mu} = s^{[kn]_m}$ , and hence an *n* and *v* such that  $N = kn + gv$ . This will be so when and only when *k* is prime to *g*.

A group may therefore be reducible to a cyclic group without itself being cyclic. It will be convenient to have the phrase "reducible to a cyclic group" cover even the irreducible cyclic groups. The class of groups reducible to cyclic groups is therefore a wider class than the class of cyclic groups. While it is obviously closed under the operation "extension of," the situation has become obscured so far as the operations "reduction of" and "subgroup of" are concerned. It turns out that the following discussion of the corresponding associated groups clears up the entire situation.

We first reinterpret *m*-adic power and *m*-adic order in terms of the coset theorem. More generally, let *s* be an element of an arbitrary *m*-group *K*,  $K^{*'}$  an arbitrary containing group of *K*. Then, in the notation of  $K^{*'}$ , the *m*-adic *n*th power  $s^{[n]}$  of *s* is the ordinary power of *s*,  $s^{(m-1)n+1}$ . The *m*-adic order of *s* is therefore the least positive integral value of *n* for which  $s^{(m-1)n+1} = s$ , i.e., for which  $s^{(m-1)n} = 1$ . It follows that the *m*-adic order *g* of *s* is identical with the ordinary order of  $s^{m-1}$ . As for the ordinary order of *s* as element of  $K^{*'}$ , we can offhand merely say that it is a divisor of  $(m-1)g$ . If, however,  $K^{*'}$  is of index *m* - 1, in particular if it be the abstract containing group  $K^*$  of *K*, then  $s^N = 1$  is possible only if *N* is a multiple of *m* - 1, and hence the ordinary order of *s* will be exactly  $(m-1)g$ .

These observations are immediately applicable to our discussion of cyclic polyadic groups, and are in turn illuminated thereby. We first observe that every containing group  $G^{*'}$  of a cyclic *m*-group *G* is cyclic. For if *s* is a generator of *G*, the elements of *G* being the *m*-adic powers of *s* are also ordinary powers of *s* in  $G^{*'}$ . Hence the elements of  $G^{*'}$ , being products of elements of *G*, are also ordinary powers of *s*, and  $G^{*'}$  is an ordinary cyclic group generated by *s*. Note that if  $G^{*'}$  is of index *m* - 1, as is always the case when *s* is an element of *K* with  $K^{*'}$  of index *m* - 1, then the order of  $G^{*'}$  is *m* - 1 times the order of *G* and thus again the ordinary order of *s*, *m* - 1 times its *m*-adic order.

Since the abstract containing group  $G^*$  of a cyclic *m*-group *G* is cyclic, it follows that its subgroup  $G_0$ , the associated ordinary group of *G*, is cyclic. Indeed, our earlier result to the effect that the *m*-adic order of *s* is equal to the



ordinary order of  $s^{m-1}$  shows that an element  $s$  of an  $m$ -group  $G$  generates  $G$  when and only when  $s^{m-1}$ , then an element of  $G_0$ , generates  $G_0$ .

This result can be immediately generalized to the following. *The associated ordinary group of a group reducible to a cyclic polyadic group is cyclic.* For the abstract containing group of the cyclic polyadic group is a containing group of the given group. The abstract associated group of the cyclic polyadic group, cyclic by the preceding result, is therefore the associated group of the given group corresponding to the above containing group. But we have shown in §6 that all containing groups of a given polyadic group yield simply isomorphic associated groups.

We have seen that every containing group of a cyclic polyadic group is cyclic. While the last argument shows that some containing group of a group reducible to a cyclic polyadic group is cyclic, it is not true that every containing group of such a group is cyclic. In fact it is readily proved that if the abstract containing group of a polyadic group is cyclic, the polyadic group itself must be cyclic. Hence, while cyclic polyadic groups are characterized by the fact that their abstract containing groups are cyclic, we must seek elsewhere for a similarly definite characterization of groups reducible to cyclic polyadic groups.

This characterization cannot consist merely of the associated ordinary group of a polyadic group being cyclic; for the abelianism of cyclic polyadic groups makes every group reducible to a cyclic polyadic group abelian, while non-abelian polyadic groups exist whose associated ordinary groups are cyclic. The added hypothesis of abelianism is however sufficient. We proceed to prove the following result which will enable us to close the entire polyadic development of cyclic groups. *Every abelian polyadic group with cyclic associated ordinary group is reducible to a cyclic polyadic group.* Since the commutativity of two elements can be tested by any extended operation, it follows that an abelian group can be reducible only to an abelian group. Coupled with the previous observations on containing and associated groups, it follows that if an abelian group with cyclic associated group is reducible to a second group, the latter is also an abelian group with cyclic associated group. Our result will therefore have been proved if we show that every irreducible polyadic group of this type is in fact cyclic.

Let then  $G$  be an irreducible abelian  $m$ -adic group of order  $g$  with cyclic associated group  $G_0$ . With  $s_0$  a fixed element of  $G$ ,  $t$  a generator of  $G_0$ , the  $g$  elements of  $G$  may be written  $s_0 t^n$ ,  $n = 0, 1, 2, \dots, g-1$ , in accordance with the coset theorem. The  $(m-1)$ -ad  $s_0^{m-1}$  will itself be in  $G_0$ . Let then  $s_0^{m-1} = t^k$ . Since  $G$  is abelian, its reducibility to a  $\mu$ -group, with  $m-1 = k(\mu-1)$ , would be equivalent to the existence of a  $(\mu-1)$ -ad  $\{s_0 t^{i_1}, s_0 t^{i_2}, \dots, s_0 t^{i_{\mu-1}}\}$  such that  $(s_0 t^{i_1} s_0 t^{i_2} \dots s_0 t^{i_{\mu-1}})^k = 1$ , i.e., such that

$$k(i_1 + i_2 + \dots + i_{\mu-1}) + \kappa \equiv 0 \pmod{g} \quad (70),$$

(70)  $G$  being abelian,  $s_0$  and  $t$  are commutative.

and hence to the H.C.F. ( $k, g$ )'s being a divisor of  $\kappa$ . It follows that the irreducibility of  $G$  is equivalent to the combined condition, each prime divisor of  $m-1$  is a divisor of  $g$ ,  $\kappa$  is prime to  $m-1$ . On the other hand, we have seen that an element  $s$  of  $G$  generates  $G$  when and only when  $s^{m-1}$  generates  $G_0$ . With  $s = s_0 t^v$ ,  $s^{m-1} = t^{v(m-1)}$ . Since, for our irreducible  $G$ ,  $\kappa$  is prime to  $m-1$ , the arithmetic progression

$$\kappa + (m-1)v, \quad v = 0, 1, 2, \dots,$$

will certainly include a value which is prime to  $g$ . With  $v$  thus chosen,  $t^{v(m-1)}$ , i.e.,  $s^{m-1}$ , is a generator of the cyclic  $G_0$ , and hence  $s$  of the consequently cyclic  $G$ .

We thus see that the class of polyadic groups reducible to cyclic polyadic groups is identical with the class of abelian polyadic groups with cyclic associated groups. The first formulation immediately showed this class of groups to be closed under the operation "extension of." The second formulation was already used to prove it closed under the operation "reduction of." It also easily shows the class to be closed under the operation "subgroup of." For such a subgroup must be abelian; while its associated group, being a subgroup of the associated group of the parent group, must be cyclic. Hence the class of polyadic groups reducible to cyclic polyadic groups is closed under the three operations "reduction of," "extension of," and "subgroup of."

In particular, the net of groups derivable from a cyclic polyadic group, or for that matter from a group reducible to a cyclic polyadic group, consists wholly of groups reducible to cyclic polyadic groups. The irreducible groups of the net are therefore all cyclic. Since we are dealing with abelian groups of finite order, the outer real dimension of these groups is 2. Hence the net of groups is in fact also derivable from an ordinary cyclic group. We proceed then to study the net of groups derivable from a cyclic 2-group of order  $g$ . Our general theory shows that for each  $m \geq 2$  there will be  $g$   $m$ -groups in the net, one for each class of equivalent  $(m-1)$ -ads of the 2-group, said class serving as the class of identities of the  $m$ -group. In terms of the given 2-group, equivalent polyads are equivalent to a unique element of the group. Hence the groups of the net are determined in 1-1 fashion by letting  $m$  run through the values 2, 3, 4,  $\dots$ , and  $s$ , the element of the 2-group equivalent to their identities, run through the  $g$  elements of that 2-group. By utilizing the expression for the operation of a polyadic group in terms of the operation of a group it is reducible to, and the fact that for any two groups of a net there is a third reducible to each, we find the following expression, in terms of the operation of the 2-group, for the operation  $c$  of an  $m$ -group of the net with identities equivalent to  $s$ :

$$c(s_1 s_2 \dots s_m) = s_1 s_2 \dots s_m s^{-1}.$$

By means of this formula we easily find which groups of the net are cyclic,

and hence also which are the irreducible groups of the net. While it also enables us to study in detail the relation of reducibility for the groups of the net, the resulting picture is quite complicated, and will not be entered into here.

Let  $s_0$  be a generator of the cyclic 2-group, and let  $s = s_0^\lambda$ . If then  $s_0^\nu$  be any element of the  $m$ -group of the net with identities equivalent to  $s$ , the above operation yields the following expression for the corresponding  $m$ -adic  $n$ th power of  $s_0^\nu$ :

$$(s_0^\nu)^{[n]} = s_0^{n(m-1)\nu + n\lambda}.$$

We then easily find the condition under which, for some  $\nu$ ,  $s_0^\nu$  is a generator of the  $m$ -group, and thus obtain the following result. *If  $s_0$  is a generator of an ordinary cyclic group of order  $g$ , the cyclic groups of the net of groups derivable from the given group are those  $m$ -groups whose identities are equivalent to an  $s_0^\lambda$  for which  $\text{H.C.F.}(m-1, \lambda, g) = 1$ . If  $\gamma$  is the order of  $s_0^\lambda$  in the 2-group, this condition is equivalent to  $g/\gamma$  prime to  $m-1$ . Thus all of the  $g$   $m$ -groups of the net for given  $m$  are cyclic when and only when  $m-1$  is prime to  $g$ . Since the irreducible groups of the net are the irreducible cyclic groups of the net, we see that the irreducible groups of the net are those for which the prime divisors of  $m-1$  are all divisors of  $g$  while  $\lambda$  is prime to  $m-1$ . Hence, for  $g \geq 2$ , a cyclic polyadic group of order  $g$  has an infinite number of outer irreducible dimensions.*

The full force of our closedness results for groups reducible to cyclic polyadic groups is brought out by the complexes obtained from such groups. We have then that the complex of groups obtainable from a cyclic polyadic group, or, in general, from a group reducible to a cyclic polyadic group, consists wholly of groups reducible to cyclic polyadic groups. We recall that the groups of any complex separate into mutually exclusive nets, there being a 1-1 correspondence between these nets and the different classes of elements the groups of the complex can have. In the present instance each net is of the type discussed above, being derivable from a group reducible to a cyclic polyadic group. Furthermore, these "group-bearing" classes now admit of very simple description. As most of the resulting picture holds good for arbitrary finite abelian polyadic groups we so present our development.

Observe first that our simplification of the operations yielding an arbitrary complex shows that its group-bearing classes, apart from that of the initial group, can all be obtained from the subgroups of the extensions of the initial group. Since a finite abelian polyadic group is always derivable from a 2-group, we may then assume that initial group to be a 2-group. That 2-group is then its own associated and containing group, and can be identified with the associated and containing group of each of its extensions. The relationship between the subgroups of a polyadic group and of its associated group, actually valid for an arbitrary containing group, then yields the following result. *The group-bearing classes of a complex obtained from a finite*

*abelian polyadic group are, apart from the class of elements of the given group, the classes of elements of the subgroups of any 2-group derivable from the given group and the cosets of those subgroups.*

We recall that the problem of the intersection of two subcomplexes of a complex was reduced to that of the intersection of their corresponding group-bearing classes. The above result then shows that, for finite abelian groups, either two group-bearing classes have no elements in common, or their common elements constitute an augmented coset of the crosscut of the subgroups of the 2-group of which they are augmented cosets. Note actually that the 2-groups derivable from the given finite abelian group are in 1-1 correspondence with the elements of the group, the element corresponding to a 2-group being the identity of the 2-group. If then  $s$  be any element of the given group, the group-bearing classes containing  $s$  constitute the subgroups of the 2-group having  $s$  as identity. It follows that *if two group-bearing classes of the complex obtained from a finite abelian group have a common element, they are the classes of elements of two subgroups of one and the same 2-group derivable from the given group, and intersect accordingly.*

In particular, then, for a cyclic polyadic group of order  $g$  the group-bearing classes of its complex are  $g/\gamma$  in number, of  $\gamma$  elements each, for every divisor  $\gamma$  of  $g$ . And two group-bearing classes, with  $\gamma_1$  and  $\gamma_2$  elements respectively, either have no elements in common, or exactly H.C.F.  $(\gamma_1, \gamma_2)$  elements in common.

The above development can be given a somewhat different turn. For any finite polyadic group a finite number of extensions of the group, and subgroups of those extensions, suffice to yield all group-bearing classes, as these are now finite in number. From the corresponding situation for a pair of groups of a net it follows that for any finite number of groups of a net there is a group of the net itself reducible to each of the given groups. Hence the above extensions can themselves be extended to one and the same group. In this process the subgroups of these groups are extended to subgroups of the resulting group. Hence, *the group-bearing classes of the complex obtained from a finite polyadic group are the classes of elements of a single suitable extension of the group, and of the subgroups of that extension.* For any finite polyadic group, therefore, the intersection of two group-bearing classes can be pictured as the intersection of two subgroups of one and the same extension of that group. And now for the earlier picture. Clearly any element of finite order in a polyadic group is of first order in some extension of that group, and hence is the sole member of a group-bearing class of the complex obtained from that group. It follows that the elements of the above "suitable extension" of a finite polyadic group are all of first order. Hence that extension will itself be reducible to each of the 2-groups derivable from the given group. If, furthermore, the given group is abelian, each of its elements  $s$  will be the identity of a 2-group to which that extension is reducible, and the subgroups of that ex-

tension containing  $s$  will thereby be reduced to the subgroups of the 2-group—hence that first picture.

In conclusion, then, while the theory of cyclic groups requires for its completion the introduction of groups reducible to cyclic polyadic groups, the theory of these groups is entirely self-contained. While it would therefore be desirable to complete this theory by developing the properties of these groups, and we have at hand the instruments that would yield this development, we have perhaps already spent too much time on such very special developments, and so pass on to the more general topics of the theory.

**23. Abstract polyadic groups of the first three orders.** The concepts of the last two sections give a certain basis for distinguishing between polyadic groups. As in ordinary theory, in counting abstract polyadic groups no distinction will be made between groups that are simply isomorphic. By contrast, in the theory of reducibility such a distinction is imperative, for two groups on the same class of elements, but with different multiplication tables, must there be considered different even if simply isomorphic. Our present interest lies not only in the results to be obtained but in the illustrations of method thus afforded.

For each  $m \geq 2$  there is of course the single abstract  $m$ -group of order one. Its sole element is of the first order, and hence the group is cyclic, and reducible to the cyclic 2-group whose sole element is the identity.

The abstract  $m$ -groups of order two can be determined directly from their possible multiplication tables<sup>(71)</sup>. If they are written on the abstract elements  $\alpha$  and  $\beta$ , and  $c$  represents the  $m$ -adic operation, the value of  $c(\alpha\alpha \cdots \alpha)$ , that is, of  $\alpha^{[1]}$ , determines the table; for each change in the value of an argument must change the value of the result. Hence there are at most two abstract  $m$ -groups of order two. It further follows that  $\alpha^{[1]}$  is, or is not, equal to  $\beta^{[1]}$  according as  $m$  is even or odd. If  $m$  is even, then if  $\alpha^{[1]} = \alpha$ ,  $\beta^{[1]} = \alpha$ , while if  $\alpha^{[1]} = \beta$ ,  $\beta^{[1]} = \beta$ , and the two possible groups are changed into each other on interchanging  $\alpha$  and  $\beta$ . On the other hand, if  $m$  is odd, if  $\alpha^{[1]} = \alpha$ ,  $\beta^{[1]} = \beta$ , and if  $\alpha^{[1]} = \beta$ ,  $\beta^{[1]} = \alpha$ , and the two groups cannot be simply isomorphic. These groups are then readily identified to yield the following result. When  $m$  is even, there is but one abstract  $m$ -group of order two, namely, the cyclic  $m$ -group of order two. It then consists of one first order element and one second order element, and is reducible to the ordinary cyclic group of order two, if it be not that group. When  $m$  is odd there are exactly two abstract  $m$ -groups of order two; one group consisting of two first order elements, and being the non-cyclic second order  $m$ -group reducible to the ordinary cyclic group of order two, the other group being the cyclic  $m$ -group of order two, consisting of two second order elements, and hence not reducible to a 2-group.

<sup>(71)</sup> Dürnte used this method to determine the number of  $m$ -groups on two symbols as elements, but did not consider the question of those  $m$ -groups being abstractly the same.



To obtain the abstract  $m$ -groups  $G$  of order three, we employ the general coset theorem method of §8. The associated ordinary group  $G_0$  must be cyclic, and hence its elements may be written  $1, t, t^2$ . If  $s_0$  be a fixed element of  $G$  with  $s_0^{m-1} = t_0, t_0$  in  $G_0$ , we may assume that either (1)  $t_0 = 1$ , (2)  $t_0 = t$ ; for were  $t_0 = t^2$ , groups simply isomorphic with those of case (2) would result.  $G_0$ , furthermore, admits of but two automorphisms, i.e., (a) the identical automorphism, (b) the automorphism interchanging  $t$  and  $t^2$  while, of course, leaving 1 invariant. With either of these automorphisms as the automorphism of  $G_0$  under  $s_0$ , and either of the two choices of  $t_0$ , an  $m$ -group will be correspondingly determined provided (A) the automorphism carries  $t_0$  into itself, (B) the  $(m-1)$ -st power of the automorphism is the automorphism of  $G_0$  under  $t_0$ . Of the four cases thus to be considered (1) (a) and (2) (a) satisfy both (A) and (B) for all  $m$ 's, and hence always determine a corresponding  $m$ -group. (1) (b) satisfies (A) for all  $m$ 's, but (B) only for  $m$  odd; for if  $m$  be even, the  $(m-1)$ -st power of the automorphism interchanges  $t$  and  $t^2$  whereas  $t_0 = 1$  leaves them unchanged. Hence (1) (b) determines an  $m$ -group when and only when  $m$  is odd. Finally, there is no polyadic group of order three corresponding to (2) (b), as (A) is then never satisfied.

We now identify and distinguish between the groups thus determined. The group (1) (a) is abelian since  $s$  and  $t$  are then commutative. Since  $G_0$  is cyclic,  $G$  is therefore cyclic, or reducible to a cyclic group. Direct calculation then shows that if  $m-1$  is a multiple of 3, each element is of first order, and hence the group is noncyclic, but reducible to the ordinary cyclic group of order three. On the other hand, when  $m-1$  is not a multiple of 3 we find that while  $s_0$  is of first order,  $ts_0$ , and in fact  $t^2s_0$ , are not, and hence must be of the third order. The group is therefore cyclic, but reducible to the ordinary cyclic group.

In the case of the group (2) (a),  $s_0$ , not being of the first order, must be of the third order. The group is therefore cyclic. When  $m-1$  is not a multiple of 3 it is therefore simply isomorphic with the group (1) (a). On the other hand, when  $m-1$  is a multiple of 3, and hence not prime to  $g=3$ , the group contains no first order element. It is therefore not reducible to an ordinary group, and consists of three third order elements.

Finally group (1) (b),  $m$  odd, is non-abelian, since  $s_0$  does not leave  $t$  invariant. Being therefore noncyclic, each of its elements is of the first order. We have already given this group with  $m=3$  as an example of one with no invariant element. This property holds for each admissible  $m$ . In fact, since any two of the three elements must generate the whole non-abelian group, each element is invariant under no other element than itself. It follows that each element transforms a second element into the third, a property which by itself can be shown to determine the multiplication table of that third order  $m$ -group for odd  $m$ . It is needless to add that this group is not reducible to an ordinary group.

The third order abstract polyadic groups may then be tabulated as follows, the numbers in the parentheses being the orders of the elements.

$$\mu=0, 1, 2, \dots$$

$m-1=$	$6\mu+1$	$6\mu+2$	$6\mu+3$	$6\mu+4$	$6\mu+5$	$6\mu+6$
cyclic (3, 3, 1)	1	1		1	1	
cyclic (3, 3, 3)			1			1
abelian (1, 1, 1)			1			1
non-abelian (1, 1, 1)		1		1		1
total	1	2	2	2	1	3

In particular, the one ordinary third order group comes under the case  $m-1=6\mu+1$  with  $\mu=0$ . We further see that the smallest value of  $m$  for which there are three abstract third order groups is 7<sup>(72)</sup>.

**24. Properties of transforms.** The coset theorem enabled us to write the transform of an element  $s$  by a polyad  $r$  in the ordinary form  $r^{-1}sr$ . A fundamental  $m$ -group is of course tacitly presupposed. Since the  $m$ -adic  $n$ th power of an element can likewise be written as an ordinary  $(m-1)n+1$  power, it follows that

$$(r^{-1}sr)^{[n]} = r^{-1}s^{[n]}r.$$

Hence, also, the  $m$ -adic order of an element is unchanged under transformation.

Suppose now that  $r^{-1}sr = s^{[a]}$ . By raising both sides of this equation to the  $m$ -adic  $\beta$ th power we then have

$$r^{-1}s^{[\beta]}r = (s^{[a]})^{[\beta]};$$

for our  $m$ -adic formula for the power of a power shows that  $(s^{[a]})^{[\beta]} = (s^{[\beta]})^{[a]}$ . Hence we have the following generalization of the corresponding ordinary theorem. *If a polyad transforms a generator of a cyclic  $m$ -group into its  $a$ th power, it transforms every element of this cyclic group into its  $a$ th power.*

Commutativity is related to transform through invariance. Given two noncommutative elements  $s_0$  and  $s$ , we consider what  $m$ -adic powers, if any, of  $s$  are commutative with  $s_0$ . If  $s$  is of  $m$ -adic order  $k$ , its ordinary order in the fundamental abstract containing group is  $(m-1)k$ . Let  $\gamma_0$  be the least positive value of  $\gamma$  for which the ordinary power  $s^\gamma$  is commutative with  $s_0$ .  $\gamma_0$  is then a divisor of  $(m-1)k$  and the distinct ordinary powers of  $s$  commutative with  $s_0$  are  $s^{n\gamma_0}$ ,  $n=1, 2, \dots, (m-1)k/\gamma_0$ . The  $m$ -adic powers  $s^{[\beta]}$  commuta-

<sup>(72)</sup> The two third order  $m$ -groups falling under the case  $m-1=6\mu+2$  have been given by Miller for  $\mu=0$

tive with  $s_0$  are those for which  $(m-1)\beta+1$  is a multiple of  $\gamma_0$ . It follows first that there will be an  $m$ -adic power of  $s$  commutative with  $s_0$  when and only when  $\gamma_0$  is prime to  $m-1$ .  $\gamma_0$  is then a divisor of  $k$ ; and if  $\beta_0$  is the least value of  $\beta$  for which  $s^{[\beta]}$  is commutative with  $s_0$ , the  $m$ -adic powers of  $s$  commutative with  $s_0$  are  $s^{[\beta_0+n\gamma_0]}$ ,  $n=0, 1, \dots, (k/\gamma_0-1)$ . Actually these  $k/\gamma_0$   $m$ -adic powers of  $s$  commutative with  $s_0$  must constitute a subgroup, necessarily cyclic, of the cyclic  $m$ -group generated by  $s$ . They are therefore the  $m$ -adic powers of some one of their number, not, however, necessarily of  $s^{[\beta_0]}$  <sup>(78)</sup>.

If we form the successive transforms

$$s^{-1}s_0s = s_1, s^{-1}s_1s = s_2, \dots, s^{-1}s_{n-1}s = s_n, \dots,$$

the resulting elements are the transforms of  $s_0$  under the various ordinary powers of  $s$ . In general, therefore, they will not all be gotten by transforming  $s_0$  by the elements of the cyclic  $m$ -group generated by  $s$ , but by the elements of the abstract containing group of that cyclic  $m$ -group, or, what is the same thing, by the various polyads of the cyclic  $m$ -group.

This suggests that given any  $m$ -group  $G$  and element  $s_0$  we consider the transforms of  $s_0$  under the various polyads of  $G$ . These will then constitute a complete set of conjugates of  $s_0$  under the abstract containing group  $G^*$  of  $G$ . The following discussion applies equally well to an  $m$ -group  $K$  taking the place of the element  $s_0$ .

With  $s$  an element in  $G$ ,  $G_0$  the associated ordinary group of  $G$ , we have the expansion  $G^* = G_0s + G_0s^2 + \dots + G_0s^{m-2} + G_0$ , with  $G_0s = G$ . Since  $G_0$  is an ordinary group, the number of transforms of  $s$  under the elements of  $G_0$  is some divisor  $\nu$  of  $g$ , the common order of  $G$  and  $G_0$ . Each coset  $G_0s^i$  therefore transforms  $s_0$  into  $\nu$  distinct elements. If two cosets yield a common transform of  $s_0$ , by writing those cosets in the form  $r_1G_0, r_2G_0$ ,  $r_1$  and  $r_2$  being elements of the cosets yielding that common transform, we see that the set of transforms yielded by one coset is identical with the set yielded by the other. The transforms of  $s_0$  under  $G^*$  thus fall into a certain number  $\kappa$  of mutually exclusive classes of  $\nu$  elements each. By a method entirely analogous to that used in the analysis of an arbitrary containing group, we easily find that  $\kappa$  is a divisor of  $m-1$ , and that the first  $\kappa$  cosets all yield distinct sets of  $\nu$  transforms each, these being repeated in order by each succeeding set of  $\kappa$  cosets. We thus have the following theorem. *The number of transforms of an element under the polyads of an  $m$ -group of order  $g$  is of the form  $\kappa\nu$ , where  $\nu$  is a divisor of  $g$ ,  $\kappa$  a divisor of  $m-1$ . For each  $i$  the  $i$ -ads of the group yield  $\nu$  distinct transforms. The  $\kappa\nu$  transforms can be obtained from the  $i$ -ads with  $i=1, 2, \dots, \kappa$ ; and these  $\kappa$  mutually exclusive sets of  $\nu$  transforms each are cyclically repeated for  $i$ -ads with  $i > \kappa$ .*

We can now connect the theory of transforms with that of groups of sub-

<sup>(78)</sup> As may be shown by an example.

stitutions. For convenience set  $\kappa = \mu - 1$ , and let  $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$  be the mutually exclusive sets of  $\nu$  transforms each corresponding to  $i = 1, 2, \dots, \mu - 1$  respectively. If  $s_i$  is any element of  $G$ , and  $s'$  is the transform of  $s_0$  by an  $i$ -ad of  $G$ ,  $s_i^{-1}s's_i$  will be the transform of  $s_0$  by an  $(i+1)$ -ad of  $G$ . It follows that  $s_i$  transforms the members of each  $\Gamma_i$  in 1-1 fashion into the members of  $\Gamma_{i+1}$ . Thus each element of  $G$  determines a  $\mu$ -adic substitution on  $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$ . Clearly, the product of  $m$  elements of  $G$  yields a  $\mu$ -adic substitution which is the product of the  $\mu$ -adic substitutions yielded by those elements. Certainly then, for our finite  $G$  the class of all  $\mu$ -adic substitutions corresponding to elements of  $G$  constitutes an  $m$ -adic group of  $\mu$ -adic substitutions isomorphic with  $G$ . It is readily seen that if  $N$  elements of  $G$  correspond to one  $\mu$ -adic substitution, exactly  $N$  elements of  $G$  correspond to each  $\mu$ -adic substitution, and the isomorphism is  $(1, N)$ . Finally, this  $(m, \mu)$  substitution group is transitive. For if element  $s'$  of  $\Gamma_i$  is the transform of  $s_0$  by the  $i$ -ad  $\{s_{j_1}, s_{j_2}, \dots, s_{j_i}\}$  of  $G$ , the transforms of  $s'$  by the elements  $s_j$  of  $G$  are the transforms of  $s_0$  by the  $(i+1)$ -ads  $\{s_{j_1}, s_{j_2}, \dots, s_{j_i}, s_j\}$  of  $G$ , hence by all  $(i+1)$ -ads of  $G$ , and so constitute the whole class  $\Gamma_{i+1}$ .

When  $\kappa = 1$ , the  $(m, \mu)$  substitution group becomes a transitive  $m$ -group of ordinary substitutions. The transforms of  $s_0$  under the elements of  $G$ , now identical with the transforms of  $s_0$  under the polyads of  $G$ , then include  $s_0$ , and are such that each is transformed into the entire set by the elements of  $G$ . On the other hand, when  $\kappa > 1$ , the transforms of  $s_0$  under the elements of  $G$  are transformed by the elements of  $G$  into an entirely different set. Nor can they then include  $s_0$ ; for  $s_0$ , being transformed into itself by that  $(m-1)$ -ad of  $G$  which is the identity of  $G$ , appears for the first time in the set of transforms for which  $i = \kappa$ . We thus see that the transforms of  $s_0$  under the elements of  $G$  must be said to constitute a complete set of conjugates of  $s_0$  under  $G$  when and only when  $\kappa = 1$ . And the fact that then and only then is  $s_0$  included in that set of transforms needs only restatement to become the following useful criterion. *The necessary and sufficient condition that the transforms of an element  $s_0$  by the elements of an  $m$ -group  $G$  constitute a complete set of conjugates under  $G$  is that  $s_0$  is commutative with some element of  $G$ .* As in the case of ordinary groups, the elements of  $G$  thus leaving  $s_0$  invariant constitute a subgroup  $H$  of  $G$ . If  $G$  is expanded in right cosets as regards  $H$ , each coset consists of all the elements of  $G$  transforming  $s_0$  into some one element. Hence, here too the number of conjugates of  $s_0$  under  $G$  is the order of  $G$  divided by the order of the largest subgroup of  $G$  leaving  $s_0$  invariant.

If  $s_0$  is actually an element of  $G$ , the above condition is automatically satisfied with  $s_0$  itself as element commutative with  $s_0$ . We thus have the significant fact that the transforms of an element of a polyadic group under the elements of the group always constitute a complete set of conjugates under the group<sup>(74)</sup>. Hence, as for ordinary groups, *all the elements of an  $m$ -group  $G$*

(74) Essentially a result of Miller's when stated for "perfect cosets."

can be separated into distinct complete sets of conjugates as regards  $G$ , and this separation can be performed in only one manner.

In the case of an  $i$ -ad of  $G$  with  $i > 1$  the transforms of the  $i$ -ad by the elements of  $G$  need no longer constitute a complete set of conjugates. Thus in the non-abelian 3-group of order three a dyad not the identity has but one transform under the elements of the group, two under the polyads of the group. However, our general theorem holds in this case; and since the  $i$ -ad is invariant under itself, it readily follows that  $\kappa$  is a divisor of  $H.C.F.(i, m-1)$ .

**25. Generation of polyadic groups by two groups, one invariant under the elements of the other.** We shall consider two distinct cases. In the first, a 2-group  $H_0$  is invariant under each element of an  $m$ -group  $K$ , in the second, an  $m$ -group  $H$  is invariant under each element of an  $m$ -group  $K$ . The discussion of the  $m$ -group  $G$  generated by the two given abstract groups can also be carried through from two different points of view, the first, that of the investigation of properties of groups assumed given, the second, that of the construction of groups hitherto unknown.

We have already illustrated the constructional point of view in §8. Our present interests being largely theoretical, we shall not further pursue the complexities introduced by that point of view in the field of abstract group theory, but merely obtain the results given by the first point of view<sup>(76)</sup>.

<sup>(76)</sup> This is the point of view really followed by Miller in §25, *Finite Groups*, despite the section heading "Construction of Groups with Invariant Subgroups." He thus obtains the theorem: "If all the elements of a group  $H$  transform  $G$  into itself, then  $H$  and  $G$  generate a group whose order is the order of  $G$  multiplied by the index under  $H$  of the crosscut of  $G$  and  $H$ ." The constructional point of view, while using his treatment for purposes of analysis, would necessitate the following complications.  $H$  and  $G$  would be given by group-satisfying multiplication tables on specified symbols as elements. These tables must then satisfy the consistency condition that  $H$  and  $G$  have at least one element in common, and that the product of two elements common to  $H$  and  $G$  is the same in  $H$  as in  $G$ . With each element of  $H$  there would be given a corresponding automorphism of  $G$  which is to be the automorphism of  $G$  induced by transforming it by that element of  $H$ . These automorphisms must then satisfy the consistency conditions that the product of the automorphisms corresponding to two elements of  $H$  is the automorphism corresponding to the product of those elements, while the automorphism corresponding to any element of  $H$  common to  $H$  and  $G$  is the automorphism of  $G$  induced by that element as element of  $G$ . That posited, our guess is that  $H$  and  $G$ , assumed finite, will generate a unique group in the sense that there exists a group  $K$  which, with respect to itself as fundamental group, is the group generated by  $H$  and  $G$ , while all such groups are simply isomorphic; a simple isomorphism being in fact determined by letting each element of  $H$  and  $G$  correspond to itself.

The above criticism assumes that we are dealing with abstract groups, the title of the chapter in which the above section appears. If the generating groups be given as substitution groups, for example, the divergence between the two points of view disappears, as there is always the symmetric group on all the letters involved to act as fundamental group. A similar situation obtains for  $m$ -adic groups of ordinary substitutions as is shown in the last footnote of our present section.

It should be pointed out that what we have termed the constructional point of view is followed in the related theory of group extensions. (See the first footnote to §8.) That it is but



This point of view assumes a given  $m$ -group  $F$ . In the first of our two cases,  $H_0$  is a subgroup of the associated ordinary group  $F_0$  of  $F$  which is invariant under each element of a subgroup  $K$  of  $F$ . It is convenient here to consider  $F$  a subgroup of itself. The crosscut of all subgroups of  $F$  which are such that  $H_0$  is a subgroup of their associated groups,  $K$  of themselves, is itself one of these subgroups, and will be said to be the  $m$ -group  $G$  generated by  $H_0$  and  $K$ . We may then also say that the  $m$ -group  $G$  generated by  $H_0$  and  $K$  is the smallest subgroup  $G$  of  $F$  such that  $H_0$  is a subgroup of  $G_0$ ,  $K$  of  $G$ . Similarly, if  $H$  and  $K$  are two subgroups of  $F$  with  $H$  invariant under each element of  $K$ , the  $m$ -group  $G$  generated by  $H$  and  $K$  is the smallest subgroup  $G$  of  $F$  such that  $H$  and  $K$  are subgroups of  $G$ . The existence and uniqueness of  $G$  is thus assured, but is entirely relative to the given  $m$ -group  $F$ .

We shall first consider the subcase of the general  $H_0, K$  case where  $K$  is the cyclic  $m$ -group generated by an element  $s$  of  $F$ . The  $m$ -group generated by  $H_0$  and  $K$  may then also be said to be generated by  $H_0$  and  $s$ . The invariance condition now reduces to  $H_0$  being transformed into itself by  $s$ . Consider the cosets  $H_0s, H_0s^{[1]}, H_0s^{[2]}, \dots$ . If  $\gamma$  is the  $m$ -adic order of  $s$ ,  $H_0s = H_0s^{[\gamma]}$ . Let then  $\kappa$  be the smallest positive integer for which  $H_0s = H_0s^{[\kappa]}$ . It then easily follows that the cosets  $H_0s, H_0s^{[1]}, \dots, H_0s^{[\kappa-1]}$  are mutually exclusive, while succeeding cosets are cyclic reproductions of these. Hence, also,  $\kappa$  is a divisor of  $\gamma$ .

We now readily show that the  $m$ -group  $G$  generated by  $H_0$  and  $s$  is given by

$$G = H_0s + H_0s^{[1]} + \dots + H_0s^{[\kappa-1]}.$$

Since  $H_0$  is a subgroup of  $F_0$ ,  $s$  an element of  $F$ , the set  $G$  thus defined is contained in  $F$ . Furthermore, the invariance of  $H_0$  under  $s$  coupled with the above coset analysis shows that the product of  $m$  elements of  $G$  is in  $G$ . Hence  $G$  is, indeed, a subgroup of  $F$ .  $G$  has  $s$  for element, in fact, as a member of  $H_0s$ . Hence  $G_0 = Gs^{-1}$  has  $H_0$  as subgroup. Finally, as with  $F$ , every subgroup of  $F$  whose associated group has  $H_0$  as subgroup, while it has  $s$  as element, contains  $G$ . Hence  $G$  is the  $m$ -group generated by  $H_0$  and  $s$ . We thus have the theorem: *If  $s$  is an element of an  $m$ -group,  $H_0$  a subgroup of the associated group of that  $m$ -group invariant under  $s$ , then if  $s^{[\kappa]}$  is the smallest positive  $m$ -adic power of  $s$  which is in the coset  $H_0s$ ,  $H_0$  and  $s$  generate an  $m$ -group whose order is  $\kappa$  times the order of  $H_0$ .*

In the general  $H_0, K$  case let  $L_0$  be the crosscut of  $H_0$  and  $K_0$ . Since  $H_0$  and  $K_0$  are invariant under each element of  $K$ , the same is true of  $L_0$ . Expand  $K$  in cosets as regards  $L_0$ , and let  $s_1, s_2, \dots, s_\kappa$  be a corresponding set of multi-

---

a related theory may be seen from the definition of an extension  $K$  of  $G$  by  $H$  as a group having  $G$  as invariant subgroup, with the quotient group  $K/G$  simply isomorphic to  $H$ . The above complication arising from the common elements of  $H$  and  $G$  goes not then arise.

pliers. We then show that the  $m$ -group  $G$  generated by  $H_0$  and  $K$  has the expansion

$$G = H_0 s_1 + H_0 s_2 + \cdots + H_0 s_k$$

with all indicated elements distinct. As a consequence of the invariance of  $H_0$  under each  $s_i$  we can reduce the product of  $m$  elements of the set  $G$  thus defined to the form  $ts$ , with  $t$  in  $H_0$ ,  $s$  in  $K$ . As  $s$  can further be written  $t's_i$  with  $t'$  in  $L_0$ , and hence in  $H_0$ ,  $s_i$  one of the above  $k$  multipliers, we see that the product in question is in  $G$ . It then follows as in the special case that  $G$  is the  $m$ -group generated by  $H_0$  and  $K$ . Moreover, suppose that with  $t_1$  and  $t_2$  in  $H_0$  we have  $t_1 s_{i_1} = t_2 s_{i_2}$ . Then  $t_2^{-1} t_1 = s_{i_2} s_{i_1}^{-1}$ . Since the left side of this equation represents an element of  $H_0$ , the right of  $K_0$ , this one element  $\tau$  is in  $L_0$ . But then  $s_{i_2} = \tau s_{i_1}$ , contradicting the assumption that  $s_1, s_2, \dots, s_k$  were the set of multipliers in question. The indicated elements of  $G$  are thus distinct, and we have the theorem: *If  $K$  is a subgroup of an  $m$ -group,  $H_0$  a subgroup of the associated group of the  $m$ -group invariant under each element of  $K$ , then  $H_0$  and  $K$  generated an  $m$ -group whose order is the order of  $H_0$  multiplied by the index under  $K$  of the crosscut of  $H_0$  and the associated ordinary group  $K_0$  of  $K$ .*

We turn now to the more interesting case of the  $m$ -group  $G$  generated by two  $m$ -groups  $H$  and  $K$ , with  $H$  invariant under each element of  $K$ . Note that  $H_0$ , the associated ordinary group of  $H$ , is then also invariant under  $K$ . It is readily seen that while the  $m$ -group generated by  $H_0$  and  $K$  is contained in the  $m$ -group generated by  $H$  and  $K$ , it will be identical with that  $m$ -group when and only when it contains an element of  $H$ . This means that for some  $t$  in  $H_0$ ,  $s$  in  $K$ ,  $s'$  in  $H$ ,  $ts = s'$ . But this is equivalent to  $s = t^{-1}s'$ , i.e., to  $s$ 's also being in  $H$ . Hence *the  $m$ -group generated by  $H$  and  $K$  is identical with the  $m$ -group generated by  $H_0$  and  $K$  when and only when  $H$  and  $K$  have a common element.*

In particular, if  $H$  and  $K$  have but one common element, while each element of  $H$  is commutative with each element of  $K$ , we shall say that the  $m$ -group  $G$  generated by  $H$  and  $K$  is their *direct product*.  $G$ , then, is also the  $m$ -group generated by  $H_0$  and  $K$ , and by  $K_0$  and  $H$ , and correspondingly has expansions which may be briefly written  $G = H_0 \times K = K_0 \times H$ . For  $L_0$  now reduces to the identity, so that in the first case, for example, the multipliers  $s_i$  are all the elements of  $K$ . More symmetrically then,  $G = (H_0 \times K_0)s$ , with  $s$ , say, the unique common element of  $H$  and  $K$ . It follows that  $G_0 = H_0 \times K_0$  is the ordinary direct product of  $H_0$  and  $K_0$ . Clearly  $s$  must be of first order, since all of its  $m$ -adic powers must be common elements of  $H$  and  $K$ ; and being invariant under  $H$  and  $K$ , it must also be invariant under  $G$ . All three groups are therefore reducible to ordinary groups, and simultaneously so. These considerations immediately extend to the "direct product" of any number of  $m$ -groups provided the unique common element is also the only element common to each group and the group generated by the remaining groups.

Special as this concept of direct product thus turns out to be, it is very useful in the theory of abstract polyadic groups. By contrast, the direct product method as applied to  $m$ -adic substitution groups, while involving no restriction on the groups per se, did not yield the  $m$ -group generated by the given  $m$ -groups, and hence is restricted in its usefulness to the construction of desired  $m$ -groups.

In the most general  $H, K$  case consider the abstract containing groups of the  $m$ -groups involved. Since  $H, K$ , and  $G$  are subgroups of the fundamental  $m$ -group  $F$ , their abstract containing groups  $H^*, K^*$ , and  $G^*$  may be considered subgroups of the abstract containing group  $F^*$  of  $F$ . As the elements of an  $m$ -group generate the corresponding containing group, it easily follows that  $G^*$  is the ordinary group generated by  $H^*$  and  $K^*$ . Since  $H^*$  will be invariant under each element of  $K^*$ , the standard theorem tells us that the order of  $G^*$  is equal to the order of  $H^*$  multiplied by the index under  $K^*$  of the crosscut of  $H^*$  and  $K^*$ . But the order of the abstract containing group of an  $m$ -group is  $m-1$  times the order of the  $m$ -group. Hence the order of  $G$  is equal to the order of  $H$  multiplied by that index.

It is easy indeed to write the actual expansion of  $G$ . According to the standard theory, if we expand  $K^*$  in cosets as regards the crosscut of  $H^*$  and  $K^*$ , and let  $r_1, r_2, \dots, r_n$  be the corresponding set of multipliers, then  $G^* = H^*r_1 + H^*r_2 + \dots + H^*r_n$  with all indicated elements distinct. If then  $r_i$  is an  $i_i$ -ad of  $K$ , the elements of  $H^*r_i$  in  $G$  will be the  $(m-i_i)$ -ads of an  $m$ -group multiplied by  $r_i$ . We may therefore write, in notation thus suggested,

$$G = (H)_{m-i_1r_1} + (H)_{m-i_2r_2} + \dots + (H)_{m-i_nr_n}.$$

Returning to the order of  $G$ , we seek a useful expression for the index in question. The crosscut  $\bar{L}$  of  $H^*$  and  $K^*$  will consist of the common  $i$ -ads of  $H$  and  $K$  for  $i=1, 2, \dots, m-1$ . For  $i=m-1$ , these common  $i$ -ads constitute the crosscut  $L_0$  of  $H_0$  and  $K_0$ . Let  $l$  be the order of  $L_0$ ,  $\kappa$  the smallest value of  $i$  for which  $H$  and  $K$  have a common  $i$ -ad. Then, by methods already made familiar, we find that  $\kappa$  is a divisor of  $m-1$ , while  $\bar{L}$  consists of  $l$   $i$ -ads for each of the  $(m-1)/\kappa$   $i$ 's,  $i=\kappa, 2\kappa, \dots, m-1$ . The order of  $\bar{L}$  is thus  $l(m-1)/\kappa$ . If now  $k$  is the order of  $K_0$ , the order of  $K^*$  is  $(m-1)k$ . Hence the index under  $K^*$  of  $\bar{L}$  is  $\kappa k/l$ . But  $k/l$  is the index under  $K_0$  of  $L_0$ . Hence the index of  $\bar{L}$  under  $K^*$  is  $\kappa$  times the index of  $L_0$  under  $K_0$ . We therefore have the following theorem. *If  $H$  and  $K$  are two subgroups of an  $m$ -group such that all the elements of  $K$  transform  $H$  into itself, then  $H$  and  $K$  generate an  $m$ -group whose order is the order of  $H$  multiplied by the index under  $K_0$  of the crosscut of  $H_0$  and  $K_0$  multiplied by a divisor  $\kappa$  of  $m-1$ , where  $\kappa$  is the smallest value of  $i$  for which  $H$  and  $K$  have a common  $i$ -ad.*

Of special interest is the case where  $K$  is the cyclic  $m$ -group generated by an element  $s$  which transforms  $H$  into itself.  $K^*$  is then an ordinary cyclic group also generated by  $s$ . Hence if  $s^\lambda$  is the smallest positive ordinary power

of  $s$  in  $H^*$ ,  $\lambda$  will be the index under  $K^*$  of the crosscut of  $H^*$  and  $K^*$ . We thus have as our first result: *If an element  $s$  of an  $m$ -group transforms a subgroup  $H$  of that  $m$ -group into itself, and if  $s^\lambda$  is the smallest positive ordinary power of  $s$  in the containing group  $H^*$  of  $H$ , then  $s$  and  $H$  generate an  $m$ -group  $G$  whose order is  $\lambda$  times the order of  $H$ . Indeed, the expansion of  $G$  is now readily seen to be*

$$G = (H)_{m-1}s + (H)_{m-2}s^2 + \cdots + (H)_{m-\lambda}s^\lambda.$$

Note that if  $k$  is the  $m$ -adic order of  $s$ , and hence  $k(m-1)$  its ordinary order,  $\lambda$  is a divisor of  $k(m-1)$ . Since the order of  $G$  must exceed the order of  $H$  whenever  $s$  is not in  $H$ , we obtain the following useful corollary further generalized below. *If  $s$  is of  $m$ -adic order one, and not in  $H$ , then the order of  $G$  is equal to the order of  $H$  multiplied by a divisor, not unity, of  $m-1$ .*

More refined results are yielded by our earlier analysis of the above mentioned index. The associated group  $K_0$  of the cyclic  $m$ -group  $K$  generated by  $s$  will be the cyclic ordinary group generated by  $s^{m-1}$ . Hence, if  $s^{\nu(m-1)}$  is the smallest positive power of  $s^{m-1}$  in  $H_0$ ,  $\nu$  will be the index under  $K_0$  of the crosscut of  $H_0$  and  $K_0$ . Consequently, *the order of  $G$  is also equal to the order of  $H$  times  $\nu$  times  $\kappa$ , where  $s^{\nu(m-1)}$  is the smallest positive power of  $s^{m-1}$  in  $H_0$ ,  $\kappa$  the smallest value of  $i$  for which  $H$  and the cyclic  $m$ -group generated by  $s$  have a common  $i$ -ad.*

We may note certain relationships between the constants thus involved. The connecting link between our two expressions for the order of  $G$  is the equation  $\lambda = \nu\kappa$ .  $\lambda$  is thus determined by  $\nu$  and  $\kappa$ . Conversely  $\nu$  and  $\kappa$  are determined by  $\lambda$  and  $m$ . For the common elements of  $H^*$  and  $K^*$  are  $s^\lambda, s^{2\lambda}, \dots, s^{k(m-1)}$ . It therefore easily follows that  $\kappa = \text{H.C.F.}(m-1, \lambda)$ , and hence  $\nu = \lambda / \text{H.C.F.}(m-1, \lambda)$ . By means of  $m$ -adic groups of ordinary substitutions it is readily shown that  $\lambda$  and  $m$  may assume arbitrary values. In the case of  $\kappa, \nu$ , and  $m$ , we have already observed that  $\kappa$  is a divisor of  $m-1$ . Our expressions for  $\kappa$  and  $\nu$  in terms of  $\lambda$  and  $m$  further show that  $(m-1)/\kappa$  is prime to  $\nu$ . Now it is readily verified that if  $\kappa, \nu$ , and  $m$  are arbitrarily chosen subject to these two conditions, then  $\lambda = \nu\kappa$  redetermines the same  $\kappa$  and  $\nu$  by means of the above formulas. It follows that  $\kappa, \nu$ , and  $m$  may assume any values subject to these conditions. If we now further introduce the  $m$ -adic orders  $h$  and  $k$  of  $H$  and  $s$ , we obtain the further conditions  $\nu$  a divisor of  $k$ ,  $h\nu$  a multiple of  $k$ ; the first from the index interpretation of  $\nu$ , the second from the order requirement imposed by  $s^{\nu(m-1)}$ 's being in  $H_0$ . We have not carried the investigation far enough, however, to see whether the resulting four necessary conditions on  $h, k, \kappa, \nu$  and  $m$  suffice to insure a corresponding  $H$  and  $s$  <sup>(7)</sup>.

<sup>(7)</sup> When  $h=k$ , the fourth condition is automatically satisfied. In this case the writer has verified by an example that  $h=k, \kappa, \nu$ , and  $m$  may have arbitrary values subject to the first three conditions.

In constructing such examples by means of  $m$ -adic groups of ordinary substitutions, we

$H$  is clearly an invariant subgroup of the generated group  $G$ . If then  $\sigma$  is the element of the  $m$ -adic quotient group  $G/H$  corresponding to  $s$ ,  $\nu$  is seen to be the  $m$ -adic order of  $\sigma$ . For the least positive  $\nu$  with  $s^{\nu(m-1)}$  in  $H_0$  is the least positive  $\nu$  with  $s^{[\nu]}$  in  $H_0s$ , and hence the least positive  $\nu$  with  $\sigma^{[\nu]} = \sigma$ . By a simple result of our later §29 the  $m$ -adic order of  $\sigma$  is a divisor of the  $m$ -adic order of  $s$ . Our previous corollary thus generalizes to the following. *The order of  $G$  is equal to the order of  $H$  multiplied by a multiple of a divisor other than unity of the  $m$ -adic order of  $s$  whenever the  $m$ -adic order of the element of  $G/H$  corresponding to  $s$  is not unity; when the latter order is unity, and yet  $s$*

are naturally led to the ordinary groups they generate as containing groups. On the other hand, our theory concerns their abstract containing groups only. In the  $\lambda, m$  example referred to above, it was possible to avoid this difficulty by so choosing  $H, K$ , and the fundamental  $F$  that their concrete containing groups were all of index  $m-1$ , and so simply isomorphic with their abstract containing groups. On the other hand, especially in the case of  $F$ , it is desirable to dispense with this requirement. For we could then fully make use of the fact that as for ordinary substitution groups, so for  $m$ -adic substitution groups, a fundamental  $F$  is always at hand, namely, the extension to an  $m$ -group of the ordinary symmetric group on all the letters involved; and clearly all fundamental  $F$ 's which are  $m$ -adic groups of ordinary substitutions yield the same  $G$ .

Actually, we can easily obtain the desired information concerning the abstract containing groups, and so the order of  $G$ , from any containing groups. We shall consider our general  $H, K$  case. Let  $F$  be a corresponding fundamental  $m$ -group,  $F^{**}$  any containing group of  $F$ . The subgroups of  $F^{**}$  generated by the elements of  $H$  and  $K$  respectively will then be containing groups  $H^{**}$  and  $K^{**}$  of  $H$  and  $K$ . In the above case of  $m$ -adic groups of ordinary substitutions,  $F^{**}$  may be the ordinary substitution group generated by the substitutions of  $F$ , in which case  $H^{**}$  and  $K^{**}$  will be the ordinary substitution groups generated by the substitutions of  $H$  and  $K$ , and so obtainable without the explicit use of  $F^{**}$ . Let  $H^{**}, K^{**}, F^{**}$  be of indices  $\rho_1, \rho_2, \rho$ . All three indices will then be divisors of  $m-1$ . Furthermore, it is readily seen that  $\rho_1$  and  $\rho_2$  will be multiples of  $\rho$ . Now if the cosets into which these containing groups are broken up are cyclically repeated until there are  $m-1$  of each, the  $i$ th cosets of  $H^{**}$  and  $K^{**}$  will be contained in the  $i$ th coset of  $F^{**}$  for  $i=1, 2, \dots, m-1$ . In particular, the  $(m-1)$ -st cosets will be the associated ordinary groups  $H'_0, K'_0, F'_0$ . And in the simple isomorphism between  $F'_0$  and  $F_0$ , the abstract associated ordinary group of  $F$ , the subgroups  $H'_0$  and  $K'_0$  will correspond to  $H_0$  and  $K_0$ . Hence, the index under  $K_0$  of the crosscut of  $H_0$  and  $K_0$  is also the index under  $K'_0$  of the crosscut of  $H'_0$  and  $K'_0$ , where the latter may now be considered the  $\rho_1$ th and  $\rho_2$ th cosets in  $H^{**}$  and  $K^{**}$ . As for  $\kappa$ , note that two products of  $i$  elements each taken from an  $m$ -group will be identical in a containing group of an  $m$ -group when and only when those two  $i$ -ads of elements are equivalent. Hence, the smallest value of  $i$  for which  $H^{**}$  and  $K^{**}$  have a common  $i$ -ad is also the smallest value of  $i$  for which  $H'_0$  and  $K'_0$ , repeated as above, have a common  $i$ -ad. Actually, the pairs of  $i$ th cosets of  $H^{**}$  and  $K^{**}$  start repeating after  $i = \text{L.C.M.}(\rho_1, \rho_2)$ . Hence,  $\kappa$  may be found from  $H^{**}$  and  $K^{**}$  if their cosets be cyclically repeated to a total number equal to  $\text{L.C.M.}(\rho_1, \rho_2)$  each. Clearly  $\kappa$  is a divisor of  $\text{L.C.M.}(\rho_1, \rho_2)$ . If desired, it is not difficult to give a number theoretic expression for  $\kappa$  in terms of the distribution of  $(i, j)$ 's for which an  $i$ -ad of  $H^{**}$  and a  $j$ -ad of  $K^{**}$  in their unpeated form are identical.

The order of  $G$  is thus determinable from  $H^{**}$  and  $K^{**}$ . Explicitly  $F^{**}$  does not enter. Hence, in the case of  $H$  and  $K$   $m$ -adic groups of ordinary substitutions, no further reference need be made to  $F$ . In particular, then, if  $H^{**}$  and  $K^{**}$  are each of index  $m-1$ , the order of  $G$  is found exactly as if they were the abstract containing groups of  $H$  and  $K$ .



is not in  $H$ , then the order of  $G$  is equal to the order of  $H$  multiplied by a divisor, not unity, of  $m-1$ .

26. *m*-adic groups of order  $g$  prime to  $m-1$ . Let  $G$  be any  $m$ -group whose order  $g$  is prime to  $m-1$ . The order of any element  $s$  of  $G$ , being a divisor of  $g$ , will then also be prime to  $m-1$ . The cyclic  $m$ -group generated by  $s$  therefore has one and only one first order element  $s_0$ , i.e.,  $s$  generates one and only one first order element  $s_0$ .  $G$  therefore has at least one first order element; and if it has exactly  $\lambda$  first order elements, all of its elements can be separated into  $\lambda$  corresponding mutually exclusive classes of elements, each class consisting of all the elements of  $G$  which separately generate the corresponding first order element. Now no first order element of  $G$  can transform another first order element of  $G$  into itself. For otherwise, by the first of the two corollaries of the last section, the two would generate a subgroup of  $G$  whose order would be a divisor, not unity, of  $m-1$ . But, as in the case of an element of  $G$ , the order of any subgroup of  $G$  must be prime to  $m-1$ . It follows that if element  $s$  of  $G$  generates the first order element  $s_0$ , and hence transforms  $s_0$  into itself, it can transform no other first order element of  $G$  into itself; for otherwise  $s_0$ , a power of  $s$ , would transform that other first order element into itself. The class of elements of  $G$  each generating  $s_0$  therefore consists of all the elements of  $G$  which transform  $s_0$  into itself, and hence constitute a subgroup of  $G$ . As this subgroup has  $s_0$  for invariant first order element, it is reducible to an ordinary group. We have thus proved that if  $G$  is an  $m$ -group whose order is prime to  $m-1$ , the elements of  $G$  can be separated into a number  $\lambda$  of mutually exclusive subgroups of  $G$ , all reducible to ordinary groups, where  $\lambda$  is the number of first order elements of  $G$ , and each subgroup contains one and only one first order element of  $G$ , and, indeed, consists of all the elements of  $G$  that transform that first order element into itself.

Other immediate consequences of the above proof are the following.  $G$  is reducible to a 2-group when and only when it has but a single first order element. If  $G$  has more than one first order element, it has no invariant element, and hence is not derivable from a 2-group. In particular, every abelian  $m$ -group whose order is prime to  $m-1$  has one and only one first order element, and hence is reducible to a 2-group<sup>(77)</sup>.

We may note in passing the marked simplicity, from the standpoint of polyadic theory, of those  $m$ -groups of order prime to  $m-1$  which are reducible to 2-groups. As seen below, the one first order element of such an  $m$ -group is also the one and only first order element of each of its subgroups. These subgroups are therefore also reducible to 2-groups. Furthermore, both the group and its subgroups are reducible to 2-groups in one and only one way. It easily follows by a slight modification of our cyclic  $m$ -group argument that when the above  $m$ -group is reduced to the 2-group, its subgroups are reduced to the subgroups of that 2-group.

<sup>(77)</sup> This generalizes a theorem of Lehmer on abelian 3-groups.

Returning to our arbitrary  $m$ -group  $G$  of order  $g$  prime to  $m-1$ , we proceed to show that the  $\lambda$  first order elements of  $G$ , as well as the corresponding  $\lambda$  subgroups into which  $G$  was decomposed, constitute a complete set of conjugates under  $G$ . It will then follow that these  $\lambda$  subgroups are all of the same order, and hence that the number of first order elements of  $G$  is a divisor of the order of  $G$ <sup>(78)</sup>. Let  $s_0', s_0'', \dots, s_0^{(\lambda)}$  be the first order elements of  $G$ ,  $k_1, k_2, \dots, k_\lambda$  the orders of the corresponding  $\lambda$  subgroups of  $G$ . Since exactly  $k_i$  elements of  $G$  transform  $s_0^{(i)}$  into itself,  $s_0^{(i)}$  is transformed into  $g/k_i$  different elements by all the elements of  $G$ . As the transform of a first order element is also of the first order,  $g/k_i \leq \lambda$ , i.e.,  $k_i \geq g/\lambda$ . Since  $g = k_1 + k_2 + \dots + k_\lambda$ , it follows that the equality sign must hold for each  $i$ . Each  $s_0^{(i)}$  therefore has the  $\lambda$  first order elements of  $G$  for its different transforms under the elements of  $G$ , whence the first half of our theorem. Now if  $s_1$  generates the first order element  $s_0^{(i)}$ , say  $s_1^{[n]} = s_0^{(i)}$ , then  $(s^{-1}s_1s)^{[n]} = s^{-1}s_1^{[n]}s = s^{-1}s_0^{(i)}s$ ; that is, the transform of  $s_1$  under  $s$  generates that first order element which is the transform of  $s_0^{(i)}$  under  $s$ . Hence, if element  $s$  of  $G$  transforms  $s_0^{(i)}$  into  $s_0^{(j)}$ , it transforms the subgroup corresponding to  $s_0^{(i)}$  into the subgroup corresponding to  $s_0^{(j)}$ , whence the rest of our result.

It follows from the above that the  $\lambda$  first order elements of  $G$  also constitute a complete set of conjugates under the  $m$ -group they generate. For that  $m$ -group will have an order prime to  $m-1$ , while its first order elements will be the  $\lambda$  first order elements of  $G$ . Since the  $m$ -group generated by a given set of elements chosen from a finite  $m$ -group will actually consist of all extended products of elements chosen from the set, it follows that the  $\lambda$  first order elements of  $G$  constitute a "generalized" complete set of conjugates under themselves, that is, each can be obtained from any other by a succession of transforms by first order elements only. Actually, this statement is weaker than the one immediately preceding, since it amounts to saying that the  $\lambda$  first order elements of  $G$  constitute a complete set of conjugates under any containing group of the  $m$ -group they generate. In any case, the question whether they constitute a complete set of conjugates under themselves, in the sense that any one can be transformed into any other by a third, is left open<sup>(79)</sup>.

We have already observed that  $\lambda$  is a divisor of  $g$ . While it is therefore prime to  $m-1$ , we now find an additional restriction imposed upon it by  $m-1$ . The first order element  $s_0'$  is of course invariant under itself. On the

<sup>(78)</sup> For, if  $k$  is the common order of these  $\lambda$  subgroups,  $g = k\lambda$ . That the number of first order elements of an arbitrary  $m$ -group need not be a divisor of its order is illustrated by the ordinary symmetric group of degree three extended to a 3-group. This 3-group of order six has four first order elements.

<sup>(79)</sup> Note that the statement of Miller, page 30 of *Finite Groups*, to the effect that the Sylow subgroups of order  $p^2$  of a group constitute a complete set of conjugates under themselves must also be interpreted in the above sense of a generalized complete set of conjugates. At least, that is all the proof there given allows us to infer.

other hand, since any other first order element  $s_0^{(0)}$  is not invariant under  $s'_0$ , it will be transformed by the polyads  $\{s'_0\}$ ,  $\{s'_0, s'_0\}$ ,  $\{s'_0, s'_0, s'_0\}$ ,  $\dots$  into a number, not unity, of first order elements which either directly, or by our general theorem on transforms, is seen to be a divisor of  $m-1$ . Since the sets of transforms of different  $s_0^{(0)}$ 's by the above polyads are mutually exclusive when not identical, a separation of the  $\lambda$  first order elements into mutually exclusive classes is thus effected, one class consisting of but one element, every other class of a number of elements which is a divisor, not unity, of  $m-1$ . Hence, if  $p_1, p_2, \dots, p_r$  are the distinct prime divisors of  $m-1$ ,  $\lambda$  is of the form  $\lambda = 1 + k_1 p_1 + k_2 p_2 + \dots + k_r p_r$ ,  $k_i \geq 0$ . In particular, if  $m-1$  is a power of a single prime  $p$ , the number of first order elements of  $G$  is of the form  $\lambda = 1 + kp$ . While for  $p > 1$  the expression for  $\lambda$  gives information concerning small  $\lambda$ 's only, every sufficiently large number being so representable, when  $m-1$  is a power of a single prime  $p$  the condition includes the condition  $\lambda$  prime to  $m-1$ , and for  $p > 2$ , is stronger than that condition.

A peculiar property of the sets of transforms arising in the preceding proof is that each set, clearly invariant under  $s'_0$ , in turn generates  $s'_0$ . More generally, any set of first order elements of  $G$  which is transformed into itself by a first order element  $s_0$  of  $G$  in turn generates  $s_0$ . This result is itself an immediate consequence of the following. A first order element of  $G$  which transforms a subgroup of  $G$  into itself must be contained in that subgroup. The proof of the last result consists in noting that, otherwise, that first order element and the subgroup would generate a subgroup of  $G$  whose order was the order of the given subgroup multiplied by a divisor, not unity, of  $m-1$ . As for the result preceding, the subgroup of  $G$  generated by the given set, being consequently invariant under  $s_0$ , must contain  $s_0$ .

Since the order of a subgroup of  $G$  must also be prime to  $m-1$ , there will be associated with every subgroup of  $G$  an existent subset of the  $\lambda$  first order elements of  $G$ , namely, the set of first order elements of the subgroup. These "group-bearing" subsets of the  $\lambda$  first order elements of  $G$  can be independently characterized as those existent subsets of the  $\lambda$  first order elements which generate no other first order elements. By the reasoning of the preceding paragraph, a first order element which transforms a group-bearing subset of first order elements into itself must be contained in that subset. As the converse must also be true, it follows that a first order element, and hence indeed any element, of  $G$  either leaves both a subgroup of  $G$  and the set of first order elements of that subgroup invariant, or else transforms neither into itself. Clearly, two subgroups of  $G$  have a common element when and only when their sets of first order elements have a common element. We finally note the following. If  $s_0^{(1)}, s_0^{(2)}, \dots, s_0^{(n)}$  are the first order elements of some subgroup of  $G$ , then of all subgroups of  $G$  with exactly those first order elements there is one contained in, and one containing each. The smallest subgroup is of course the crosscut of all the subgroups in question, and will indeed be

the subgroup  $H$  generated by those first order elements<sup>(80)</sup>. Now let  $K$  be the subgroup of  $G$  consisting of all the elements of  $G$  which transform  $H$  into itself.  $K$  will then contain all of the above subgroups. And since each of the first order elements of  $K$  transforms  $H$  into itself, they will all be in  $H$ , and hence will be the given first order elements.  $K$  is therefore that largest subgroup of our theorem.

The above theory is significant only if there exist  $m$ -groups of order prime to  $m-1$  with more than one first order element, and, preferably, not consisting wholly of first order elements. For odd  $m-1 \neq 1$  such an  $m$ -group is furnished by the complete  $m$ -adic  $\delta$ -group which is of order  $2^{m-1}$  and has  $2^{m-2}$  first order elements. The  $2^{m-2}$  second order subgroups are then the corresponding mutually exclusive subgroups into which the elements of the group are separated. For  $m-1$  even, and  $\lambda$  prime to  $m-1$ , the  $\lambda$  second order elements of the ordinary dihedral group of order  $2\lambda$  constitute such an  $m$ -group under the product of  $m$  elements as operation. In this  $m$ -group all  $\lambda$  elements are of  $m$ -adic order one. However, by the direct product method, we can obtain from this  $m$ -group, and a cyclic  $m$ -group of order  $g/\lambda$ , an  $m$ -group of arbitrary order  $g$  prime to the even  $m-1$ , and with an arbitrary divisor  $\lambda$  of  $g$  as the number of its first order elements. Most of the theory can be illustrated by means of these examples.

27. **Sylow subgroups of order  $p^a$  with  $g/p^a$  prime to  $m-1$ .** That Sylow's theorem is not universally valid for polyadic groups is shown by cyclic polyadic groups. We recall that a cyclic  $m$ -group of order  $g$  has a subgroup of order  $\gamma$ ,  $\gamma$  a divisor of  $g$ , when and only when  $g/\gamma$  is prime to  $m-1$ . Hence, if  $p$  is a prime divisor of  $g$ , and  $p^a$  is the largest power of  $p$  which divides  $g$ , a cyclic  $m$ -group of order  $g$  will have a "Sylow subgroup" of order  $p^a$  when and only when  $g/p^a$  is prime to  $m-1$ . This example shows that our extension of Sylow's theorem to polyadic groups as given below is the most general that can be given in terms of a condition involving only the order and dimension of the group<sup>(81)</sup>. Note also that our cyclic group will have a Sylow subgroup for each of two distinct prime divisors of  $g$  when and only when  $g$  itself is prime to  $m-1$ , in which case it will have a Sylow subgroup for every distinct

<sup>(80)</sup> In this connection a theorem of Dörnte's is of interest. To wit, if an  $m$ -group is semi-abelian, and has at least one first order element, then its first order elements themselves constitute a subgroup of the  $m$ -group.

<sup>(81)</sup> Other theorems however are possible. Thus, if  $G$  is an  $m$ -group of order  $g$  whose associated ordinary group  $G_0$  has but one Sylow subgroup corresponding to a prime divisor  $p$  of  $g$ , in particular if  $G$  is semi-abelian, then the necessary and sufficient condition that  $G$  have a Sylow subgroup corresponding to  $p$  is that  $G$  have at least one element whose order is a power, possibly the zeroth, of  $p$ . Necessary, immediately; and sufficient. For if  $H_0$  is that sole Sylow subgroup of  $G_0$  of order a power of  $p$ ,  $s$  the element of  $G$ , then  $s$  can transform  $H_0$  only into itself, while  $s^{m-1}$ , being of ordinary order a power of  $p$ , must be in  $H_0$ . Hence  $H = H_0 s$  is an  $m$ -group, and thus a subgroup of  $G$  of the requisite order. However, the Sylow subgroups of  $G$  corresponding to the prime  $p$  need not then constitute a complete set of conjugates under  $G$ . Thus, if  $G'$  is

prime divisor of  $g$ . The same situation holds for the applicability of our extension of Sylow's theorem to polyadic groups.

We proceed then to prove the following. *If the order  $g$  of an  $m$ -group  $G$  is divisible by  $p^a$  but not by  $p^{a+1}$ ,  $p$  a prime divisor of  $g$ , then if  $g/p^a$  is prime to  $m-1$ ,  $G$  will have at least one subgroup of order  $p^a$ .* Our proof consists in expressing  $G$  in accordance with our basic coset theorem, and applying the Sylow theorem for ordinary groups to the associated ordinary group  $G_0$  of  $G$ . By that coset theorem, and in the notation of the abstract containing group  $G^*$  of  $G$ , we may write  $G = s'G_0$ , where  $s'$  is any element of  $G$ . Since  $G_0$  is also of order  $g$ , it will have at least one Sylow subgroup  $H_0$  of order  $p^a$ . As  $G_0$  is invariant under  $s'$ ,  $H_0$  will be transformed by  $s'$  into a Sylow subgroup  $H'_0$  of  $G_0$  of order  $p^a$ . But the Sylow subgroups of  $G_0$  of order  $p^a$  constitute a complete set of conjugates under  $G_0$ . Hence some element  $t$  of  $G_0$  will transform  $H'_0$  into  $H_0$ . It follows that the element  $s'' = s't$  of  $G$  transforms  $H_0$  into itself.

Now  $s''$  as element of  $G$  will be of some  $m$ -adic order  $\gamma$  which is a divisor of  $g$ . If then  $p^\beta$  is the largest power of  $p$  which divides  $\gamma$ ,  $\gamma/p^\beta$  will be prime to  $m-1$ . It follows from our theory of cyclic groups that  $s''$  will generate an element  $s$ , also in  $G$ , of  $m$ -adic order  $p^\beta$ . That is,  $s$  as element of  $G^*$  will be of ordinary order  $p^\beta(m-1)$ , and hence  $s^{m-1}$  of ordinary order  $p^\beta$ . But  $H_0$ , being invariant under  $s''$ , must also be invariant under  $s$ , and hence under  $s^{m-1}$ . Since  $s^{m-1}$  of order  $p^\beta$  is in  $G_0$ , and transforms Sylow subgroup  $H_0$  of  $G_0$  of order  $p^a$  into itself,  $s^{m-1}$  must be in  $H_0$ . It follows from the converse of the coset theorem that  $H = H_0s$  is an  $m$ -group, hence a subgroup of  $G$ , and of order  $p^a$ .

Our proof actually shows then that for each Sylow subgroup of order  $p^a$  of  $G_0$  there is at least one "Sylow subgroup" of order  $p^a$  of  $G$  whose associated ordinary group is that Sylow subgroup of  $G_0$ . Conversely, the associated ordinary group of any subgroup of order  $p^a$  of  $G$  will be a subgroup of order  $p^a$  of  $G_0$ , and hence a Sylow subgroup of order  $p^a$  of  $G_0$ . Since one and only one subgroup of  $G_0$  can be the associated ordinary group of a given subgroup of  $G$ , we thus see that there is a one-many correspondence thus set up between the Sylow subgroups of order  $p^a$  of  $G_0$ , and those of  $G$ .

Of the three results which together constitute Sylow's theorem for ordi-

---

an ordinary abelian group, some extension of it  $G$ , also abelian, will consist wholly of first order elements. There will then be  $g/p^a$  Sylow subgroups of  $G$  of order  $p^a$ , yet each is invariant under  $G$ .

Again, in attempting to generalize the standard substitution group proof of the existence of Sylow subgroups by means of  $m$ -adic substitution groups, the writer succeeded in constructing a Sylow subgroup corresponding to the prime  $p$  for any symmetric  $m$ -adic substitution group of degree a power of  $p$ . It may be of interest to note that the rest of that standard proof goes over except for the last step. This one point of failure, and failure there must be for an arbitrary  $m$ -group, lay in our being able to establish that the number of elements in a double coset  $H_1sH_2$  was the order of a subgroup of  $H_1$  only for the case when  $H_1$  and the transform of  $H_1$  under  $s$  have a common element.



nary groups we have therefore proved that the first, pertaining to the existence of Sylow subgroups, go over for polyadic groups under the given order condition. We now show that under the same condition the third result also goes over. That is, *under the condition of the preceding theorem the Sylow subgroups of order  $p^a$  of the  $m$ -group  $G$  constitute a complete set of conjugates under  $G$ .* We have to show then that each subgroup of order  $p^a$  of  $G$  can be transformed into any other by an element of  $G$ . Let  $H'$  and  $H$  be any two such Sylow subgroups of  $G$ ,  $H'_0$  and  $H_0$  the corresponding Sylow subgroups of  $G_0$ . Some element  $t$  of  $G_0$  will transform  $H'_0$  into  $H_0$ . That same  $t$  will then transform  $H'$  into a Sylow subgroup  $H''$  of  $G$  also corresponding to  $H_0$ , i.e., having  $H_0$  for associated ordinary group. If then we can show that some element  $s'$  of  $G$  will transform  $H''$  into  $H$ , it will follow that element  $s = ts'$  of  $G$  must transform  $H'$  into  $H$  as required by our theorem.

Our problem therefore reduces to showing that of all Sylow subgroups  $H^{(i)}$  of  $G$  corresponding to one and the same Sylow subgroup  $H_0$  of  $G$ , each can be transformed into any other by an element of  $G$ . Since  $H_0$  is the associated ordinary group of each  $H^{(i)}$ , it will be transformed into itself by the elements of each  $H^{(i)}$ . If then  $\bar{G}$  is the subgroup of  $G$  consisting of all the elements of  $G$  which transform  $H_0$  into itself, each  $H^{(i)}$  will be a subgroup of  $\bar{G}$ . On the one hand, therefore, Lagrange's theorem for polyadic groups shows that if  $\bar{g}$  is the order of  $\bar{G}$ , then  $\bar{g}$  will be divisible by  $p^a$ , but not by  $p^{a+1}$ , while  $\bar{g}/p^a$  will be prime to  $m-1$ . On the other hand, since  $H_0$  is invariant under each element of  $\bar{G}$ , it will be an invariant subgroup of  $\bar{G}_0$ , the associated ordinary group of  $\bar{G}$ . First then,  $H_0$ , whose order proclaims it to be a Sylow subgroup of  $\bar{G}_0$ , is the only Sylow subgroup of  $\bar{G}_0$  of order  $p^a$ . And since  $\bar{G}$  satisfies the order condition of our first theorem, it follows from the proof of that theorem that the subgroups  $H^{(i)}$ , which constitute all the Sylow subgroups of order  $p^a$  of  $G$ , and hence of  $\bar{G}$ , corresponding to  $H_0$ , actually are the only subgroups of order  $p^a$  of  $\bar{G}$ .

If we expand  $\bar{G}$  in cosets as regards  $H_0$ , each subgroup  $H^{(i)}$ , having  $H_0$  for associated group, will appear as one of these cosets. Since  $H_0$  is invariant under each element of  $\bar{G}$ , these cosets are the elements of the  $m$ -adic quotient group  $\Gamma = G/H_0$ .  $H_0$  then appears as the identity of  $\Gamma_0$ , the associated ordinary group of  $\Gamma$ , each  $H^{(i)}$  as an element  $\sigma^{(i)}$  of  $\Gamma$ . If  $s$  is an element of  $H^{(i)}$ ,  $s^{m-1}$  is in  $H_0$ . Hence for each  $\sigma^{(i)}$ ,  $[\sigma^{(i)}]^{m-1} = 1$ . That is, each  $\sigma^{(i)}$  is a first order element of the  $m$ -group  $\Gamma$ . Conversely, if  $\sigma$  be any first order element of  $\Gamma$ , the corresponding coset of  $\bar{G}$  constitutes a subgroup of  $\bar{G}$  with  $H_0$  for associated group, and hence is an  $H^{(i)}$ . The elements  $\sigma^{(i)}$  are therefore the only first order elements of  $\Gamma$ . But the order of  $\Gamma$  is  $\bar{g}/p^a$  which is prime to  $m-1$ . The preceding section therefore tells us that the elements  $\sigma^{(i)}$  constitute a complete set of conjugates under the elements of  $\Gamma$ . It follows that each of the subgroups  $H^{(i)}$  of  $\bar{G}$  can be transformed into any other by an element of  $\bar{G}$ , and hence of  $G$ . Our proof is thus completed.

Clearly, the Sylow subgroups of order  $p^a$  of  $G$  are also the Sylow subgroups of order  $p^a$  of the subgroup of  $G$  generated by those Sylow subgroups. As that generated subgroup must satisfy the order condition of our theorem, it follows that the Sylow subgroups of order  $p^a$  also constitute a complete set of conjugates under the elements of the  $m$ -group they generate. As in the case of the preceding section, a weaker form of this result is that the Sylow subgroups of order  $p^a$  of  $G$  constitute a generalized complete set of conjugates under their own elements, that is, each can be obtained from another by a succession of transforms by their own elements.

Under the condition  $g/p^a$  prime to  $m-1$ , two of the three parts of Sylow's theorem have thus been shown to hold verbatim for polyadic groups. Not so for the remaining part concerning the number of Sylow subgroups of order  $p^a$ . Let us return to the one-many correspondence between the Sylow subgroups of order  $p^a$  of  $G_0$  and of  $G$ . As stated in different guise in the preceding proof, an element  $t$  of  $G_0$  which transforms one Sylow subgroup of  $G_0$  into a second will transform the Sylow subgroups of  $G$  corresponding to that first Sylow subgroup of  $G_0$  into those corresponding to the second. Each Sylow subgroup of order  $p^a$  of  $G_0$  therefore has the same number  $\lambda$  of corresponding Sylow subgroups of  $G$ . As seen above,  $\lambda$  is actually the number of first order elements of an  $m$ -group of order  $g/p^a$  prime to  $m-1$ . Hence our result of the preceding section, coupled with the corresponding part of the Sylow theorem for ordinary groups, yields the following as the remaining part of our Sylow theorem for polyadic groups. *Under the condition of the preceding theorems the number of Sylow subgroups of order  $p^a$  of the  $m$ -group  $G$  of order  $g$  is of the form  $(1+kp)\lambda$  where  $\lambda$  is a divisor of  $g/p^a$  and hence prime to  $m-1$  and  $p$ .*

In contrast with the above, we are able to extend the ordinary result that every element and subgroup of order a power of  $p$  is contained in a Sylow subgroup of order  $p^a$ , only for several still narrower classes of polyadic groups. It will be convenient to refer to this as the *inclusion property*. We do have immediately that *under the conditions of the preceding theorems if element  $s$  of order  $p^\beta$  of  $G$ ,  $\beta \geq 0$ , transforms a Sylow subgroup  $H$  of order  $p^a$  of  $G$  into itself, then  $s$  is in  $H$* . For otherwise, by our generalized corollary of §25,  $s$  and  $H$  would generate a subgroup of  $G$  whose order would be either  $p^a$  times a multiple of  $p$ , or  $p^a$  times a divisor, not unity, of  $m-1$ , neither of which possibility is consistent with the given conditions. Hence also, if each element of a subgroup  $K$  of order  $p^\beta$  of  $G$  transforms  $H$  into itself, then  $K$  is contained in  $H$ . It follows that if  $G$  has but one Sylow subgroup of order  $p^a$ , in particular then if  $G$  is abelian, the inclusion property holds. Again, as in the proof of the first part of our extension of Sylow's theorem, we see that if element  $s$  of order  $p^\beta$  of  $G$  transforms a Sylow subgroup  $H_0$  of order  $p^a$  of the associated ordinary group  $G_0$  into itself, then  $s$  must be in a Sylow subgroup of order  $p^a$  of  $G$ , namely,  $H_0$ ; likewise then for a subgroup  $K$  of order  $p^\beta$  of  $G$  that transforms  $H_0$  into itself. For  $K_0$  will then be contained in  $H_0$ ; and with  $s$  in  $K$ , Sylow

subgroup  $H_0s$  of  $G$  will contain  $K = K_0s$ . Hence, if  $G_0$  has but one Sylow subgroup, in particular if  $G_0$  is abelian, i.e.,  $G$  semi-abelian, the inclusion property is satisfied.

If we attempt to generalize the standard proof of the inclusion property for ordinary groups, we see that while the number of Sylow subgroups of order  $p^\alpha$  of the  $m$ -group  $G$  is shown by our formula to be again prime to  $p$ , our work on transforms merely shows the number of transforms of a Sylow subgroup under the polyads formed from  $s$  or  $K$  to be a divisor of  $p^\beta(m-1)$ . We are thus led to the inclusion property only when  $m-1$  itself is a power of the prime  $p$ . More generally, however, let  $G$  be reducible to a  $\mu$ -group  $G'$ , with  $\mu-1$  a power of  $p$ , say  $p^\gamma$ . The abstract containing group  $G'^*$  of  $G'$ , of order  $p^\gamma g$ , will then be a containing group of  $G$ . The corresponding containing group of the cyclic  $m$ -group generated by  $s$ , or of  $K$ , will be a subgroup of  $G'^*$ . It follows that the above number of transforms will also be a divisor of  $p^\gamma g$ , and hence actually be a power of  $p$ . The standard proof therefore again generalizes. Hence, *under the condition of the preceding theorems the inclusion property holds whenever  $G$  is reducible to a  $\mu$ -group with  $\mu-1$  a power of  $p$ ; in particular, then, whenever  $G$  is reducible to an ordinary group.*

An interesting consequence of this result is that the inclusion property for  $G$  holds under the condition of this section *whenever  $G$  has an invariant element*. For let  $s$  be an invariant element of  $G$ . Since its  $m$ -adic order is a divisor of  $g$ , the condition  $g/p^\alpha$  prime to  $m-1$ , coupled with our formula for the real dimension of a cyclic  $m$ -group, shows that the cyclic  $m$ -group generated by  $s$  is reducible to a  $\mu$ -group with  $\mu-1$  a power of  $p$ . If then we apply our general criterion of reducibility to a  $\mu$ -group to this cyclic  $\mu$ -group, we obtain a condition which, with the invariance of  $s$  under  $G$ , becomes the condition that  $G$  be reducible to a  $\mu$ -group. Note that in this case, which is that of a  $G$  derivable from a 2-group, for each Sylow subgroup of order  $p^\alpha$  of  $G_0$  there is but one corresponding Sylow subgroup of  $G$ . For the invariant element  $s$  will generate some invariant element of order a power of  $p$ , which, consequently, must be in every Sylow subgroup of order  $p^\alpha$  of  $G$ . On the other hand two Sylow subgroups of  $G$  corresponding to the same Sylow subgroup of  $G_0$  can have no common element.

All of the above concerned the Sylow subgroups of  $G$  corresponding to the single prime  $p$ . As stated early in this section, if the condition  $g/p^\alpha$  prime to  $m-1$  is to be satisfied for two distinct prime factors of  $g$ , then  $g$  itself must be prime to  $m-1$ , in which case the condition is satisfied for every prime factor of  $g$ . Hence, when  $g$  is prime to  $m-1$ , our extension of Sylow's theorem is universally valid. In particular, if  $G$  is abelian with  $g$  prime to  $m-1$ , then  $G$  has one and only one Sylow subgroup for each distinct prime divisor of  $g$ . By the preceding section,  $G$  then has one and only one first order element, which must then be in each of the Sylow subgroups of  $G$ , and, indeed, be the only element common to one such subgroup and the subgroup generated by

the others.  $G$ , therefore, is then the direct product of its Sylow subgroups; and when it is reduced to a 2-group, in the one manner allowed by its unique first order element, its Sylow subgroups are reduced to the Sylow subgroups of that 2-group.

Actually, this last result is but a special instance of a general result. We have earlier observed that when an  $m$ -group  $G$  is reduced to a  $\mu$ -group  $G'$ , each subgroup of  $G'$  is the reduction of a subgroup of  $G$ , but a subgroup of  $G$  may not reduce to a subgroup of  $G'$ . On the other hand, let  $G$  satisfy our general condition  $g/p^a$  prime to  $m-1$ . Then  $G'$  satisfies the corresponding condition  $g/p^a$  prime to  $\mu-1$ . Our extension of Sylow's theorem is therefore applicable to both groups. Since transforms of elements by elements are the same in  $G$  and in  $G'$ , our complete set of conjugates result, applied to a Sylow subgroup of order  $p^a$  of  $G'$  and that of  $G$  reducing to it, shows that *when  $G$  is reduced to  $G'$  the Sylow subgroups of order  $p^a$  of  $G$  are reduced to the Sylow subgroups of order  $p^a$  of  $G'$* . Finally, if  $m-1$  is prime to  $g$ , the Sylow subgroups of  $G$ , without qualification, are reduced to the Sylow subgroups of  $G'$ .

**28. Representation of an arbitrary  $m$ -adic group as a regular  $m$ -adic substitution group.** We shall prove our result without the use of the coset theorem. The proof will then, indeed, immediately lead to another proof of the coset theorem, actually, the writer's original proof<sup>(2)</sup>.

Let  $G$  be an arbitrary  $m$ -group of order  $g$ . The classes  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  are then to have for members the  $g$  classes of equivalent  $i$ -ads for  $i=1, 2, \dots, m-1$ . It will be convenient to symbolize the  $g$  members of  $\Gamma_i$  by  $a_{ij}, j=1, 2, \dots, g$ . Let  $s$  be any element of  $G$ . Then, as proved in more general form in §3, if the  $i$ -ads  $\{s'_1, s'_2, \dots, s'_i\}$  and  $\{s''_1, s''_2, \dots, s''_i\}$  of  $G$  are equivalent, the  $(i+1)$ -ads  $\{s'_1, s'_2, \dots, s'_i, s\}$  and  $\{s''_1, s''_2, \dots, s''_i, s\}$  of  $G$  are equivalent, and conversely.  $s$  thus becomes an operator which carries the  $g$  classes of equivalent  $i$ -ads in 1-1 fashion into the  $g$  classes of equivalent  $(i+1)$ -ads. Furthermore, if  $c$  represents the  $m$ -adic operation of  $G$ , then if the  $(m-1)$ -ads  $\{s'_1, s'_2, \dots, s'_{m-1}\}$  and  $\{s''_1, s''_2, \dots, s''_{m-1}\}$  are equivalent, the elements  $c(s'_1 s'_2 \dots s'_{m-1} s)$  and  $c(s''_1 s''_2 \dots s''_{m-1} s)$  are identical, and conversely. It follows that  $s$  thus carries in 1-1 fashion the letters of  $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$ , that is, determines an  $m$ -adic substitution on the  $\Gamma$ 's.

Now given any  $i$ -ad  $\{s_1, s_2, \dots, s_i\}$ , and any  $(i+1)$ -ad  $\{s'_1, s'_2, \dots, s'_i, s'_{i+1}\}$ , there is one and only one element  $s$  of  $G$  for which the  $(i+1)$ -ads  $\{s_1, s_2, \dots, s_i, s\}$  and  $\{s'_1, s'_2, \dots, s'_i, s'_{i+1}\}$  are equivalent. It follows on the one hand that no two distinct elements of  $G$  can yield the same  $m$ -adic substitution on the  $\Gamma$ 's. The correspondence between the elements of  $G$  and the  $m$ -adic substitutions they determine is therefore 1-1. And since the  $m$ -adic substitution determined by  $c(s_1 s_2 \dots s_m)$  is clearly the product of the  $m$ -adic substitutions determined by  $s_1, s_2, \dots, s_m$ , it follows that the  $m$ -adic substitu-

<sup>(2)</sup> While the proof as given is for finite  $m$ -groups, it holds with little change for all  $m$ -groups. Hence the full generality of the consequent proof of the coset theorem.

tions determined by the elements of  $G$  constitute an  $m$ -adic substitution group simply isomorphic with  $G$ . Furthermore, the initial observation of this paragraph shows that given any two letters in successive  $\Gamma$ 's there is one and only one element  $s$  of  $G$ , and hence one and only one  $m$ -adic substitution of the simply isomorphic substitution group, that carries the letter in the first  $\Gamma$  into that of the second. This  $m$ -adic substitution group is therefore regular. We have consequently proved the following generalization of Cayley's theorem. *Every  $m$ -adic group can be represented as a regular  $m$ -adic substitution group.* In this connection, as seen in §16, the argument of §14 shows that *two regular  $m$ -adic substitution groups on the same letters which are simply isomorphic are conjugate.*

If we now wish to obtain the coset theorem from this result, we need merely observe that the ordinary group generated by the  $m$ -adic substitutions of the representation of  $G$ , as in the case of all  $m$ -adic substitution groups, is a containing group of the representation of  $G$  of index  $m-1$ , and hence by resymbolization of its elements can be made a containing group of  $G$  leading to the desired result. Since we have developed our theory of abstract polyadic groups abstractly, comparatively few applications of this generalization of Cayley's theorem are to be found in the present paper. Perhaps the most important of these is that it allows the concept of holomorph to apply to an arbitrary abstract polyadic group.

**29. Invariant subgroups and quotient groups; the  $m$ -adic central quotient group.** The present section may be considered a continuation of §4, our attention now being restricted to finite polyadic groups. We recall that if  $G$  is an  $m$ -group with ordinary associated group  $G_0$ , then every subgroup  $H_0$  of  $G_0$  that is invariant under  $G$  leads to an  $m$ -adic quotient group  $Q=G/H_0$  isomorphic with  $G$ . Clearly, if  $H_0$  is of order  $h$ , the isomorphism between  $G$  and  $Q$  is  $(h, 1)$ .  $H_0$  and  $Q$  may be called complementary groups as regards  $G$ . Since the elements of  $Q$  are the cosets of  $G$  as regards  $H_0$ , the order of  $G$  is the product of the orders of  $H_0$  and  $Q$ . Similarly for an actual subgroup  $H$  of  $G$  corresponding to  $H_0$ .

Let  $\sigma$  be any element of  $Q$ ,  $s$  any one of the elements of the corresponding coset. Then the  $m$ -adic order  $n$  of  $s$  must be divisible by the  $m$ -adic order  $\nu$  of  $\sigma$ . For, since  $s^{[n]}=s$ ,  $\sigma^{[n]}=\sigma$ , and hence  $n$  is a multiple of  $\nu$ . That is, *the order of any element of an  $m$ -adic quotient group divides the orders of all the elements of the corresponding coset.* We recall that each coset corresponding to a first order element of  $Q$  constitutes a subgroup of  $G$ . These subgroups in fact are all the subgroups of  $G$  having  $H_0$  for associated ordinary group, and hence also are semi-invariant subgroups of  $G$ . In particular, if  $H_0$  is of order prime to  $m-1$ , each coset thus corresponding to a first order element of  $Q$  has at least one first order element.

Unlike the corresponding situation for ordinary groups, an element  $\sigma$  of  $Q$  may be of order a power of a prime  $p$  without any element of the correspond-



ing coset being of order a power of that prime. Thus, let  $G$  be a cyclic  $m$ -group of order  $p^\alpha k$  where  $k$ , prime to  $p$ , is not prime to  $m-1$ . Then no element of  $G$  can have an order a power of  $p$ . But with  $H_0$  the subgroup of  $G_0$  of order  $k$ ,  $Q = G/H_0$  is cyclic, and of order  $p^\alpha$ . Some element  $\sigma$  of  $Q$  will then indeed be of order  $p^\alpha$ , while the corresponding coset has no element of order a power of  $p$ .

However, let  $\sigma$  be of order  $p^\beta$ ,  $H_0$  of order  $p^\alpha k$ ,  $k$  prime to  $p$ , and suppose that  $k$  is prime to  $m-1$ . The elements of the cosets corresponding to the  $m$ -adic powers of  $\sigma$  will then together constitute a subgroup  $G'$  of  $G$  of order  $p^{\alpha+\beta}k$ . Since  $k$  is prime to  $m-1$ ,  $G'$  will have a Sylow subgroup  $K$  of order  $p^{\alpha+\beta}$ <sup>(88)</sup>. As the crosscut of  $K_0$  and  $H_0$  must be of order a power of  $p$ , it follows that  $K$  must have exactly  $p^\alpha$  elements in each of the  $p^\beta$  cosets of  $G'$  as regards  $H_0$ . The coset corresponding to  $\sigma$  therefore has at least one element of order  $p^\gamma$  with, of course,  $\gamma \geq \beta$ . That is, *if the order of an element of an  $m$ -adic quotient group is a power of a prime number  $p$ , while the largest divisor prime to  $p$  of the order of the complementary group is prime to  $m-1$ , then the corresponding coset involves an element whose order is a power of  $p$ .*

We recall the ordinary group result that every invariant subgroup of index 2 under any group includes all the elements of odd order contained in this group. In the case of an  $m$ -adic quotient group of order two, we recall our results of §23, and note that for  $m$  odd no such result can be expected. In fact, when the quotient group consists of two first order elements, each of the corresponding cosets, both then invariant subgroups of the given group as a consequence of the abelianism of the quotient group, may have an element of odd order; while when the quotient group consists of two second order elements both cosets consist of even order elements only. On the other hand, for  $m$  even the quotient group must consist of one first and one second order element. The coset corresponding to the first order element of the quotient group will then be an invariant subgroup of the given group, and any elements of odd order in the given group must be included in that invariant subgroup.

If  $H_0$  is a subgroup of  $G_0$ , the index of  $H_0$  under  $G$  may be defined as the order of  $G$  divided by the order of  $H_0$ , and, of course, gives the number of cosets in the expansion of  $G$  in either right or left cosets as regards  $H_0$ —likewise for an  $H$  actually a subgroup of  $G$ . In the case of ordinary groups, we know that the index of the crosscut of two subgroups of a group under one of those subgroups is less than or equal to the index of the other subgroup under the group; while if the two subgroups are conjugate under the group, the inequality always prevails. If now  $H$  is a subgroup of an  $m$ -group  $G$ ,  $K_0$  a subgroup of  $G_0$ , let  $L_0$  be the crosscut of the associated ordinary group  $H_0$  of  $H$ , and  $K_0$ . Then, by writing  $G$  in the form  $G_0s$ , with  $s$  in  $H$ , we see that the expansion of  $H_0$  in right cosets as regards  $L_0$ , and the expansion of  $G_0$  in right

<sup>(88)</sup> Unless  $\alpha = \beta = 0$ . But that case has already been treated. Actually, the first order elements of  $G$  may then conveniently be considered its Sylow subgroups of order  $p^0$ .

cosets as regards  $K_0$ , become the expansions of  $H$  and  $G$  in right cosets as regards  $L_0$  and  $K_0$  respectively. It then follows immediately that the index of  $L_0$  under  $H$  is less than or equal to the index of  $K_0$  under  $G$ . Now let  $K_0$  be the associated ordinary group of a subgroup  $K$  of  $G$  conjugate to  $H$  under  $G$ . Since  $H$  and  $K$  are subgroups of  $G$ , we see from the discussion in §24 that  $H$  can also be transformed into  $K$  by some element  $t$  of  $G_0$ . Since  $t$  then transforms  $H_0$  into  $K_0$ , the 2-group result for conjugate subgroups is applicable and thus yields the following. *If  $H$  and  $K$  are conjugate subgroups of an  $m$ -group  $G$ , the index of the crosscut of  $H_0$  and  $K_0$  under one of the subgroups is always less than the index of these subgroups under  $G$ .*

In this formulation we use "subgroup" in the strict sense, and thereby avoid the need of specifying that  $H_0$  and  $K_0$ , or  $H$  and  $K$ , are distinct. Now, as in the corresponding 2-group illustration, let  $H$  be of index 2 under  $G$ . With  $K$  conjugate to  $H$ , the above result shows  $H_0$  and  $K_0$  to be identical.  $H$  is then at least a semi-invariant subgroup of  $G$ . But since the resulting quotient group  $G/H$ , being of order two, is abelian, it follows that  $H$  is actually invariant under  $G$ . Hence, as for ordinary groups, a subgroup of index 2 under any polyadic group is invariant.

If an  $m$ -group  $G$  has at least one invariant element, these invariant elements clearly constitute an invariant subgroup of  $G$  which may be called the central of  $G$ . Note that a necessary and sufficient condition that our finite  $m$ -group  $G$  have a central is that it be derivable from an ordinary group. The central  $C$  of  $G$ , when it exists, is of course abelian, and coincides with  $G$  when and only when  $G$  is abelian. The quotient group  $G/C$  may be called the central quotient group of  $G$ , and, as with ordinary groups, is easily proved noncyclic whenever  $G$  is non-abelian.

It is readily seen that, when the central  $C$  of  $G$  exists, the associated ordinary group  $C_0$  of  $C$  consists of all the elements of  $G_0$  which are invariant under  $G$ . In general then, let us define the associated central  $C_0$  of  $G$  as the subgroup of  $G_0$  consisting of all the elements of  $G_0$  invariant under  $G$ .  $C_0$  then always exists, and being a subgroup of  $G_0$  invariant under  $G$ , always leads to a quotient group  $G/C_0$ . Since  $G/C = G/C_0$  whenever  $C$  exists, we may call  $G/C_0$  the central quotient group of  $G$  irrespective of the existence of  $C$ . Since each element of  $C_0$  is also invariant under  $G_0$ ,  $C_0$  is a subgroup of the central of  $G_0$  when it does not coincide with the central of  $G_0$ . It is readily seen, in fact, that the central of  $G_0$  is invariant under  $G$ , each element of  $G$  yielding the same automorphism of that central. It follows that  $C_0$  consists of those elements of the central of  $G_0$  which are left invariant under any one element of  $G$ . In particular, when  $C$  exists,  $C_0$  will coincide with the central of  $G_0$ . In any case,  $C_0$  is abelian, and coincides with  $G_0$  when and only when  $G$  is abelian. It is then again easily proved that the central quotient group of an  $m$ -group  $G$  is noncyclic whenever  $G$  is non-abelian.

Any subgroup of  $G$  having  $C_0$  for associated group leads to the central

quotient group  $G/C_0$  and may be called a *relative central* of  $G$ . The relative centrals of  $G$  are then those cosets, if any, of the expansion of  $G$  as regards  $C_0$  which correspond to first order elements of the central quotient group. They are of course semi-invariant subgroups of  $G$ , and are easily seen to be abelian. They can be independently characterized as the maximal subgroups of  $G$  having the property that, on being transformed by an element of  $G$ , each element of the subgroup is multiplied by one and the same element  $t$  of  $G_0$ . Together, the elements of the relative centrals of  $G$  constitute all elements  $s$  of  $G$  with  $s^{m-1}$  in  $C_0$ . The relative centrals corresponding to invariant first order elements of the central quotient group are characterized by the above multiplier  $t$ 's always being in  $C_0$ , in which case, indeed,  $t^{m-1} = 1$ . The unique central  $C$ , when it exists, is then the only one for which  $t$  is always 1.

30. **Commutator, semi-commutator, and quasi-commutator subgroups.** A direct extension to polyadic groups of the concepts of commutator, and commutator subgroup, is immediately obtainable. Given an  $m$ -group  $G$ , and in the notation of the abstract containing group of  $G$ , if  $s_1$  and  $s_2$  are any two elements of  $G$ , we may, as in ordinary theory, define the commutator of  $s_1$  and  $s_2$  to be  $t = s_1^{-1}s_2^{-1}s_1s_2$ . We shall also refer to  $s_1$  and  $s_2$  as the elements of the commutator. The commutator of  $s_1$  and  $s_2$  is then not an element of  $G$ , but of  $G_0$ , the associated ordinary group of  $G$ , and is indeed that element of  $G_0$  by which  $s_1$  has to be multiplied on the right to yield the transform of  $s_1$  under  $s_2$ . The different commutators thus formed from elements of  $G$  therefore generate a subgroup of  $G_0$ , if not  $G_0$  itself, which may then be called the *commutator subgroup* for  $G$ .

As in ordinary group theory, the theory of commutator subgroups for polyadic groups is intimately bound up with the property of abelianism. But now our general formulation of semi-abelianism given in §7 suggests the need of a corresponding formulation of semi-commutator subgroup. The relative complexity of the resulting formulation then suggests a still further generalization of both concepts to what we term quasi-abelianism, and quasi-commutator subgroup. This wider generalization is also significant for ordinary groups. But while thus intimately related to certain recent work, in particular of Hall and Neumann<sup>(84)</sup>, its direction seems to be new.

The immediate connection between abelianism and commutator subgroup is more clearly in evidence if we rewrite the usual  $s_1s_2 = s_2s_1$  for the former in the equivalent form  $s_1^{-1}s_2^{-1}s_1s_2 = 1$ . Now the expression  $s_1^{-1}s_2^{-1}s_1s_2$  that thus enters into both concepts is but a special instance of a word in the sense of Hall, or a rational expression in the sense of Baer. In general, a word  $W$  will be any expression of the form  $s_1^{a_1}s_2^{a_2} \cdots s_n^{a_n}$ , where the exponents are arbitrarily  $+1$  or  $-1$ , the subscripts arbitrarily equal or unequal. If such an expression is to assume the value 1 for any choice of  $s$ 's in an  $m$ -group  $G$ , the notation

<sup>(84)</sup> B. H. Neumann, *Identical relations in groups* I, Mathematische Annalen, vol. 114 (1937), pp. 506-525. References will here be found to the work of Hall.

being that of the abstract containing group of  $G$ , the exponents must satisfy the condition  $\nu_1 + \nu_2 + \cdots + \nu_N \equiv 0 \pmod{m-1}$ . Given  $m$ , consider then any specific class of words  $W_i$  whose exponents satisfy this condition. An  $m$ -group  $G$  will then be said to be *quasi-abelian* of corresponding formal type if the equations  $W_i = 1$  are satisfied for every assignment of elements in  $G$  as values of the  $s$ 's, i.e., form a set of identical relations for  $G$  in the sense of Neumann. Now given an arbitrary  $m$ -group  $G$ , as a result of the exponent condition on the given class of words  $W_i$  each word assumes an element of  $G_0$  as value when its letters are assigned elements of  $G$  as values. We shall call these words formal quasi-commutators, their values quasi-commutators, of the given formal type. The subgroup of  $G_0$  generated by all of the quasi-commutators thus obtainable from elements of  $G$  will then be called the *quasi-commutator subgroup* for  $G$  of corresponding formal type.

In particular, any formulation of semi-abelianism as given in §7 can be rewritten in the above form. We correspondingly have formal semi-commutators, semi-commutators, and *semi-commutator subgroup* for an  $m$ -group  $G$ . While a certain degree of arbitrariness enters into the manner in which the equations of §7 are thus rewritten, it will be seen that this is irrelevant in the formation of the corresponding semi-commutator subgroup for  $G$ . In fact, our central theorem will be to the effect that the correspondence between type of quasi-abelianism and type of quasi-commutator subgroup, at present purely formal, is in fact intrinsic<sup>(85)</sup>.

Our initial development, paralleling that of ordinary theory up to its main conclusion, will be given for quasi-commutator subgroups, the results then also holding for the successive specialization to semi-commutator and commutator subgroups. Consider then any one formulation of quasi-commutator subgroup for  $m$ -groups. From its very definition we then have that *the quasi-commutator subgroup for an  $m$ -group  $G$  reduces to the identity when and only when  $G$  is quasi-abelian of corresponding formal type*. Clearly the transform  $W_i$  by  $s$  is the same expression with each letter in  $W_i$  replaced by its transform

(<sup>85</sup>) Note that while we are interested in all, in the present instance finite,  $m$ -groups satisfying a given set of identical relations, Neumann considered instead the class of all identical relations satisfied by a given, of course ordinary, group. But it is the former concept that generalizes abelianism. Again, Hall, in the first paper cited by Neumann, builds up higher commutator forms merely out of ordinary commutators. His later concept of word-subgroup is identical, for ordinary groups, with our quasi-commutator subgroup. But again the emphasis is on all word-subgroups of a given group, rather than word-subgroup of given type for all groups—say of cardinal number less than, or less than or equal to, a given cardinal. And so our particular contribution of the relation between type of word-subgroup and type of identical relations is again unnoticed. We hasten to add that the researches of these authors in the directions they do pursue are profound. We also note that on reading Neumann's paper we changed our original formulation involving a finite number of identical relations to an arbitrary set of identical relations. In the case of our formulation of semi-abelianism, the finite can stand; for our theorem of §7 shows that an infinite set would always be equivalent to a finite subset thereof.

under  $s$ . That is, the transform of each quasi-commutator by an element of  $G$  is also a quasi-commutator. Hence, the *quasi-commutator subgroup* for  $G$  of the given formal type is a subgroup of  $G_0$  invariant under  $G$ , when not  $G_0$  itself. We may therefore form the  $m$ -adic quotient group of  $G$  relative to this quasi-commutator subgroup, i.e., the corresponding *quasi-commutator quotient group* of  $G$ . We then readily see that as in the ordinary theory, the *quasi-commutator quotient group* of  $G$  of given formal type is quasi-abelian of the corresponding formal type. For the isomorphism between  $G$  and the quotient group shows that a quasi-commutator formed from any elements of the quotient group corresponds to the quasi-commutator formed in the same way from corresponding elements of  $G$ , and hence is always the identity. Conversely, consider any quotient group of  $G$  which is quasi-abelian according to the given formulation. Again quasi-commutators of  $G$  correspond to quasi-commutators of this quotient group. Since the latter quasi-commutators can only be the identity, the former must be in the subgroup of  $G_0$  complementary to this quotient group. That is, every subgroup of  $G_0$  which is invariant under  $G$ , and whose complementary quotient group is quasi-abelian of given formal type, contains the *quasi-commutator subgroup* for  $G$  of corresponding formal type.

We are now able to prove the following fundamental theorem. *If two formulations of quasi-abelianism for  $m$ -adic groups are such that every  $m$ -group satisfying either satisfies the other, then the corresponding quasi-commutator subgroups for an  $m$ -group are always identical.* For let  $A'$  and  $A''$  symbolize the two formulations of quasi-abelianism. If then, for a given  $m$ -group  $G$ ,  $C'_0$  and  $C''_0$  are the quasi-commutator subgroups corresponding to  $A'$  and  $A''$  respectively, the quasi-commutator quotient group  $G/C'_0$  satisfies  $A'$ ,  $G/C''_0$  satisfies  $A''$ . By our hypothesis, therefore, the  $m$ -group  $G/C'_0$  also satisfies  $A''$ ,  $G/C''_0$  also satisfies  $A'$ . Hence, by our last theorem,  $C'_0$  contains  $C''_0$  and  $C''_0$  contains  $C'_0$ , that is,  $C'_0$  and  $C''_0$  are identical.

The converse of this theorem is immediate; for if two formulations of quasi-commutator subgroup lead to identical subgroups for each  $m$ -group, then, if either of these subgroups is the identity, the other also is the identity. If then we say that two formulations of quasi-abelianism for  $m$ -adic groups define the same *type of quasi-abelianism* if every  $m$ -group satisfying either satisfies the other, while two formulations of quasi-commutator subgroup for  $m$ -adic groups define the same *type of quasi-commutator subgroup* if they yield identical subgroups for each  $m$ -group, we can conclude that there is a 1-1 correspondence between types of quasi-abelianism for  $m$ -adic groups and types of quasi-commutator subgroup. The correspondence between quasi-abelianism and quasi-commutator subgroup, originally depending on a particular formulation, has thus been shown to be intrinsic.

A useful partial consequence of our earlier proof is the following. *If two formulations of quasi-abelianism for  $m$ -adic groups are such that every  $m$ -group satisfying the first satisfies the second, then the quasi-commutator subgroup for an*



*m*-group corresponding to the first formulation always contains the one corresponding to the second. In this connection note that quasi-commutator subgroups of different types may be identical for a particular *m*-group. We therefore pause to prove the following. Given any finite set of distinct types of quasi-abelianism, there exists an *m*-group for which the corresponding quasi-commutator subgroups are all distinct. In fact, for each pair of these types there must exist an *m*-group quasi-abelian according to one type, but not according to the other. Represent these *m*-groups say as *m*-adic substitution groups on different letters, and form the *m*-group *G* therefrom by the direct product method. *G* then has the desired property. For it is readily proved from commutativity considerations that each quasi-commutator of *G* is the product of quasi-commutators of the same form, one for each of the above constituent groups of *G*, and conversely. Hence the quasi-commutator subgroups for *G* corresponding to any two of the given types of quasi-abelianism have, on the letters of the corresponding constituent group of *G*, a constituent group which is the identity in one case, not the identity in the other, and hence are themselves distinct.

Our basic "equivalence theorem" immediately translates our determination of the distinct types of semi-abelianism effected in §7 into a determination of the distinct types of semi-commutator subgroup. Since the proof of distinctness for the former was carried through by means of finite groups, we can therefore state that *there are as many distinct types of semi-commutator subgroups for m-adic groups as there are distinct divisors of m-1*. For a divisor  $\rho$  of  $m-1$ , the semi-commutator subgroup corresponding to  $\rho$ -semi-abelianism may be called the  $\rho$ -semi-commutator subgroup. From the above more general result it follows that *there exists an m-group for which the semi-commutator subgroups of all the distinct types are distinct*. In this case a simpler example of such a group is obtained merely by taking the direct product of groups, one for each divisor  $\rho-1$  of  $m-1$ , which, as in §7, are *m*-groups  $\rho$ -semi-abelian, but not  $\rho'$ -semi-abelian for any divisor  $\rho'-1$  of  $\rho-1$  other than  $\rho-1$ . Whether the semi-commutator subgroups of a given *m*-group are distinct or not, we may note the following relations between them. Since  $\rho_1$ -semi-abelianism implies  $\rho_2$ -semi-abelianism whenever  $\rho_1-1$  is a divisor of  $\rho_2-1$ , it follows that in this case the  $\rho_1$ -semi-commutator subgroup contains the  $\rho_2$ -semi-commutator subgroup. More generally then, the crosscut of the  $\rho_1$  and  $\rho_2$ -semi-commutator subgroups contains the  $\rho_3$ -semi-commutator subgroup, where  $\rho_3-1 = \text{L.C.M.}(\rho_1-1, \rho_2-1)$ , while the subgroup generated by the  $\rho_1$  and  $\rho_2$ -semi-commutator subgroups is contained in the  $\rho$ -semi-commutator subgroup, where  $\rho-1 = \text{H.C.F.}(\rho_1-1, \rho_2-1)$ . In the second case, however, we can prove that *the subgroup generated by the  $\rho_1$  and  $\rho_2$ -semi-commutator subgroups is the  $\rho$ -semi-commutator subgroup with  $\rho-1 = \text{H.C.F.}(\rho_1-1, \rho_2-1)$* . For by the general theorem of §7, the semi-abelianism defined by the combination of  $\rho_1$ -semi-abelianism and  $\rho_2$ -semi-abelianism is equivalent to  $\rho$ -semi-

abelianism with the above  $\rho$ . The  $\rho$ -semi-commutator subgroup is therefore also the subgroup generated by all semi-commutators of the  $\rho_1$  and  $\rho_2$  formal types, and hence by the  $\rho_1$  and  $\rho_2$ -semi-commutator subgroups themselves.

In our march to the equivalence theorem we neglected certain developments related only to semi-commutators, or merely commutators, which might well have come first. In the limited generality of the first specialization we note that *each semi-commutator subgroup for an  $m$ -group  $G$  contains the commutator subgroup of the ordinary associated group  $G_0$  of  $G$* . In fact, if  $H_0$  be such a semi-commutator subgroup, the quotient group  $G_0/H_0$  can be identified as the associated ordinary group of the semi-commutator quotient group  $G/H_0$ . Since  $G/H_0$  is semi-abelian,  $G_0/H_0$ , by a result of §7, is abelian, whence the above.

Clearly, two elements of a polyadic group are commutative when and only when their commutator is the identity. As in the corresponding situation for ordinary groups, it is readily proved that if the elements of a commutator respectively belong to two invariant subgroups of a polyadic group, the commutator is contained in the crosscut of the associated ordinary groups of those subgroups. It follows that *if two invariant subgroups of a polyadic group are such that their associated ordinary groups have only the identity in common, then every element of one of these subgroups is commutative with every element of the other*. Since two subgroups having at least one element in common have as many elements in common as have their associated ordinary groups, the above result is in this case equivalent to the following. *If two invariant subgroups of a polyadic group have one and only one element in common, then every element of one of these subgroups is commutative with every element of the other*. Actually, this special case is almost an immediate consequence of the corresponding ordinary theorem; for the one common element is then an invariant first order element of each of the subgroups, and hence of the polyadic group they generate<sup>(\*)</sup>, so that all three of these groups are reducible, and simultaneously so, to ordinary groups.

We have observed that the commutator of elements  $s_1$  and  $s_2$  of  $G$  is the element of  $G_0$  which must be multiplied into  $s_1$  to obtain the transform of  $s_1$  under  $s_2$ . Hence the complete set of conjugates of  $s_1$  under  $G$  can be obtained by multiplying  $s_1$  by commutators formed from elements of  $G$ . Since the commutator subgroup for  $G$  is invariant under  $G$ , it readily follows from this that all the transforms of an  $i$ -ad of  $G$  by polyads of  $G$  can be obtained by multiplying the  $i$ -ad by elements of the commutator subgroup for  $G$ . More specifically, it can be proved by way of the equivalence theorem that the transforms of an  $i$ -ad of an  $m$ -group  $G$  by the elements of  $G$  can be obtained by multiplying one such transform by elements of the  $\rho$ -semi-commutator subgroup for  $G$ , where  $\rho - 1 = \text{H.C.F.}(i, m - 1)$ ; whence likewise for the transforms of the  $i$ -ad by the  $j$ -ads of  $G$  with fixed  $j$ . It follows from this result that if  $G$  is  $\rho$ -semi-

(\*) Their direct product, therefore, as defined in §25.

abelian, all elements of  $G$  transform the  $i$ -ad into the same  $i$ -ad, as also do all  $j$ -ads with fixed  $j$ , a fact also easily shown directly.

We have defined an  $m$ -group  $G$  to be simple if  $G_0$  has no subgroup other than the identity invariant under  $G$ . It follows then immediately that if a simple  $m$ -group  $G$  is not quasi-abelian of specified type, the corresponding quasi-commutator subgroup for  $G$  is identical with  $G_0$ . If then, rather narrowly, we define  $G$  to be perfect if the commutator subgroup for  $G$  is identical with  $G_0$ , it follows that every simple polyadic group of composite order is perfect. For otherwise  $G$  would be abelian, while  $G_0$  would possess a subgroup other than the identity, yet invariant under  $G$ .

As in the case of ordinary groups, a subgroup of an  $m$ -group  $G$  may be called a characteristic subgroup of  $G$  if it corresponds to itself under every automorphism of  $G$ . Every automorphism of  $G$  determines an automorphism of  $G_0$ . We may then define a subgroup of  $G_0$  to be an associated characteristic subgroup of  $G$  if it corresponds to itself under every automorphism of  $G$ . In the case of invariance, a subgroup of  $G_0$  invariant under  $G$  is always invariant under  $G_0$ , but not conversely. Here the reverse situation holds. For clearly a characteristic subgroup of  $G_0$  is also an associated characteristic subgroup of  $G$ , but not always conversely, as shown by the following example. The complete  $m$ -adic  $\delta$ -group for  $m=3$  is a triadic group of order four which has exactly two second order subgroups, one cyclic, the other non-cyclic. Each of the subgroups is therefore a characteristic subgroup of the group. Evidently the associated ordinary group of any characteristic subgroup of a polyadic group is an associated characteristic subgroup of the group. On the other hand, the associated ordinary group of this triadic  $\delta$ -group is the ordinary axial group, and hence itself has no characteristic subgroup of order two.

It is readily proved that if  $G$  is non-abelian, then the central of  $G$ , if existent, is a characteristic subgroup of  $G$ , while the associated central of  $G$  is an associated characteristic subgroup of  $G$ . We now observe that every quasi-commutator subgroup for  $G$ , when not identical with  $G_0$ , is an associated characteristic subgroup of  $G$ . In fact it is readily seen that under any automorphism of  $G$  a quasi-commutator involving certain elements of  $G$  will correspond to a quasi-commutator of the same form involving the corresponding elements of  $G$ . As the first set of elements take on all values in  $G$ , so do the second, so that actually the set of quasi-commutators of  $G$  of given formal type corresponds to itself under the automorphism.

Granting that the concept of quasi-abelianism and quasi-commutator subgroup has a certain degree of generality, ever further generalizations suggest themselves<sup>(87)</sup>. Perhaps a guiding principle in such generalizations might

(87) Thus, if the above concepts be termed categorical, the following generalization, which we give only for ordinary groups, can be effected. With each of a given class of words  $W_j$  is associated a class of words  $W_{j\pm}$  involving only the letters of  $W_j$ . A group  $G$  will then be conditionally quasi-abelian of corresponding formal type if each  $W_j = 1$  is satisfied for every assign-

be the existence of an equivalence theorem. It may then be of interest to present our equivalence theorem in the following light. Each type of quasi-commutator subgroup for  $m$ -groups may be thought of as a function which assumes for each  $m$ -group  $G$  a subgroup of  $G_0$ , if not  $G_0$ , as value. Our equivalence theorem then asserts that this function is completely determined when it is known for what values of its argument it assumes the value 1.

**31. The  $\phi$ -subgroup of an  $m$ -adic group.** The concept of a set of elements of a group being a set of independent generators of the group is equally applicable to a polyadic group. Whereas an ordinary group always has at least one element, namely the identity, which can never be one of a set of independent generators of the group<sup>(88)</sup>, this need not be so in the case of a polyadic group. Thus a cyclic  $m$ -group of order  $g$  such that each prime divisor of  $g$  divides  $m-1$  can be generated by any one of its elements, and hence fails to possess an element of the type in question. If, however, an  $m$ -group  $G$  has at least one element which cannot be one of a set of independent generators of the group, then the set of all such elements constitutes a characteristic subgroup of  $G$  which may be called the  $\phi$ -subgroup of  $G$ . It is a mark of the generality of the concept of the  $\phi$ -subgroup that the self-same proofs which yield the corresponding results for ordinary groups apply verbatim to polyadic groups to give the following. *The  $\phi$ -subgroup of an  $m$ -group  $G$  is the crosscut of all the maximal subgroups of  $G$ . If the  $\phi$ -subgroup of an  $m$ -group  $G$  involves a non-invariant element or subgroup, the number of conjugates under  $G$  of this element or subgroup is greater than the number of the corresponding conjugates under the  $\phi$ -subgroup.* As an application of the first of these results we may note that if a cyclic  $m$ -group  $G$  is of order  $g = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_p^{\alpha_p} \gamma_0$ , the  $p$ 's being the distinct prime divisors of  $g$  not divisors of  $m-1$ , then the  $\phi$ -subgroup of  $G$  exists if there be at least one such prime  $p$ , and is then the subgroup of order  $p_1^{\alpha_1-1} p_2^{\alpha_2-1} \cdots p_p^{\alpha_p-1} \gamma_0$ . Hence also, if we continue forming  $\phi$ -subgroups starting with the cyclic  $m$ -group  $G$ , we finally arrive at the subgroup of order  $\gamma_0$  which has no  $\phi$ -subgroup. Since the  $\phi$ -subgroup is always a "proper" subgroup, if we start with any finite  $m$ -group and successively form  $\phi$ -subgroups,

ment of elements in  $G$  as values of its letters for which each  $W_{jk} = 1$  is satisfied. Correspondingly, the conditional quasi-commutator subgroup of  $G$  is to be the smallest subgroup of  $G$  having the property that each  $W_j$  is in that subgroup for every assignment of elements in  $G$  as values of its letters for which each  $W_{jk}$  is in that subgroup. Our development up to, and including, the equivalence theorem then goes over. But now symbolic logic suggests that our conditions might involve more explicitly its apparent variables and other apparatus, and our horizon keeps receding. Thus, also, Neumann suggests the possibility of allowing constant elements of a group to enter into his identical relations, while Hall, in his higher commutator forms, from the start allows arbitrary subgroups of  $G$  individually to replace  $G$  as domain of a corresponding variable. It may be that a postulational procedure, perhaps centering around our actual development, or around the point of view about to be suggested, would bring order out of the chaos that thus threatens.

(88) Unless the group is the identity.

we arrive at a subgroup whose  $\phi$ -subgroup is nonexistent. This weak statement, supported by the above example for  $m > 2$ , contrasts with the case  $m = 2$  when the last existent  $\phi$ -subgroup is always the identity.

In applying the second of the above two general results to the Sylow subgroups of the  $\phi$ -subgroup of an arbitrary  $m$ -group  $G$ , we are hampered by the order condition of our extension of Sylow's theorem. Within the scope of that condition, we note first that if the  $\phi$ -subgroup of  $G$  is of order  $g'$ , and if, with  $p^{a'}$  the largest power of the prime  $p$  dividing  $g'$ ,  $g'/p^{a'}$  is prime to  $m-1$ , then the  $\phi$ -subgroup has a Sylow subgroup of order  $p^{a'}$  which then, as in the ordinary case, is unique. If then  $g'$  itself is prime to  $m-1$ , the  $\phi$ -subgroup will have one and only one Sylow subgroup for each distinct prime divisor of  $g'$ . Since, with  $g'$  prime to  $m-1$ , the first order elements of the  $\phi$ -subgroup constitute a complete set of conjugates under the  $\phi$ -subgroup, it follows as for the Sylow subgroups that the  $\phi$ -subgroup then has one and only one first order element. That is, when the order of the  $\phi$ -subgroup of an  $m$ -group is prime to  $m-1$ , the  $\phi$ -subgroup is reducible to a 2-group. When so reduced its Sylow subgroups are reduced to the Sylow subgroups of the 2-group. As in the ordinary case, the  $\phi$ -subgroup is then the direct product of its Sylow subgroups.

This result has an interesting consequence when the order of the given  $m$ -group is itself prime to  $m-1$ . The  $\phi$ -subgroup, if it exists, then has but one first order element. The invariance of the  $\phi$ -subgroup therefore entails the invariance of this first order element under the given  $m$ -group. But this can only be the case if the  $m$ -group has no other first order element. Hence, *if an  $m$ -group of order prime to  $m-1$  has more than one first order element, its  $\phi$ -subgroup is nonexistent*; that is, if an  $m$ -group of order prime to  $m-1$  is not reducible to a 2-group, each of its elements can be one of a set of independent generators of the group. On the other hand, if the  $m$ -group is reducible to a 2-group, its sole first order element can be generated by any other element, and hence is in the consequently existent  $\phi$ -subgroup of the group.

We restrict our discussion of the  $\phi$ -subgroups of primitive groups to primitive  $m$ -adic groups of ordinary substitutions. By the corresponding theorem of §18, the subgroups consisting of all substitutions omitting a given letter are maximal subgroups. Since these maximal subgroups can only have the identity in common, it follows that *the  $\phi$ -subgroup of a primitive  $m$ -adic group of ordinary substitutions is either the identity, or else is nonexistent*. Certainly then when the primitive group in question does not possess the identity, and hence a fortiori when it is not reducible to a 2-group, its  $\phi$ -subgroup is nonexistent. Strangely enough, the same may be true even when the identity is in the primitive group, then consequently reducible to a 2-group. Thus, the ordinary cyclic substitution group of order and degree a prime  $p$  remains primitive when extended to a  $(p+1)$ -group. Yet, while the identity and any other element together generate the  $(p+1)$ -group, each alone generates only itself.



32. **Simply isomorphic  $m$ -adic groups; group of inner isomorphisms.** We have defined simply isomorphic  $m$ -groups in §4, and have shown there that the transform of an  $m$ -group by an element or polyad is an  $m$ -group simply isomorphic with the given  $m$ -group. Restricting our attention to the case when the simple isomorphism is an automorphism, i.e., between an  $m$ -group and itself, we then have conversely, as in the case of ordinary groups, that any automorphism of an  $m$ -group can be effected by transforming it by an element. This really means that an  $m$ -group can be found of which the given  $m$ -group is a subgroup and which has an element so transforming the given  $m$ -group. This result may be proved as in the ordinary case by representing the given  $m$ -group as a regular  $m$ -adic substitution group in accordance with §28. Then, by §16, the principal holomorph of the  $m$ -group so represented certainly transforms it into each of its possible automorphisms.

Since the abstract containing group of an  $m$ -group is determined abstractly by the  $m$ -group, we see that a simple isomorphism between two  $m$ -groups determines a simple isomorphism between their abstract containing groups. Conversely, any simple isomorphism between the abstract containing groups of two  $m$ -groups which makes the classes of elements of the  $m$ -groups correspond determines a simple isomorphism between the  $m$ -groups. The simple isomorphism theorem of §8 may be considered a refinement of this obvious result. As that theorem is related to the determination theorem preceding it, so the following theorems are related to two of the generation theorems of §25. Their proofs, easily supplied, are therefore here omitted.

*Two  $m$ -groups of the same order  $G'$  and  $G''$  are simply isomorphic if their associated ordinary groups  $G'_0$  and  $G''_0$  contain two simply isomorphic subgroups  $H'_0$  and  $H''_0$  invariant under  $G'$  and  $G''$  respectively, while  $G'$  and  $G''$  are generated by  $H'_0$  and  $H''_0$  and two elements  $s_1$  and  $s_2$  such that if  $s_1^{(m-1)}$  is the smallest positive power of  $s_1^{m-1}$  that occurs in  $H'_0$ , then  $s_2^{(m-1)}$  is the smallest positive power of  $s_2^{m-1}$  that occurs in  $H''_0$ , and  $s_1^{(m-1)}$ ,  $s_2^{(m-1)}$  correspond in the given simple isomorphism of  $H'_0$  and  $H''_0$ . Moreover, it is assumed that  $s_1$  and  $s_2$  transform corresponding generators of  $H'_0$ ,  $H''_0$  into corresponding elements in the given simple isomorphism.*

*Two  $m$ -groups of the same order  $G_1$  and  $G_2$  are simply isomorphic if they contain two simply isomorphic invariant subgroups  $H_1$  and  $H_2$  respectively, and are generated by these subgroups and two elements  $s_1$  and  $s_2$  such that if  $s_1^1$  is the smallest positive power of  $s_1$  which occurs in the abstract containing group  $H_1^*$  of  $H_1$ , then  $s_2^1$  is the smallest positive power of  $s_2$  which occurs in the abstract containing group  $H_2^*$  of  $H_2$ , and  $s_1^1$  and  $s_2^1$  correspond as a consequence of the given simple isomorphism of  $H_1$  and  $H_2$ . Moreover, it is assumed that  $s_1$ ,  $s_2$  transform corresponding generators of  $H_1$ ,  $H_2$  into corresponding elements in the given simple isomorphism.*

We have observed that cyclic  $m$ -groups of the same order are simply isomorphic, and, obviously, no noncyclic  $m$ -group can be simply isomorphic

with a cyclic  $m$ -group. The following is a rather interesting application of the simple isomorphism theorem of §8. Let  $G'$  and  $G''$  be two  $m$ -groups of order  $g$  reducible to cyclic polyadic groups, and let element  $s'_0$  of  $G'$  be of the same  $m$ -adic order as element  $s''_0$  of  $G''$ . Then element  $s'^{m-1}_0$  of  $G'_0$  is of the same ordinary order as element  $s''^{m-1}_0$  of  $G''_0$ . Since  $G'_0$  and  $G''_0$  are ordinary cyclic groups of order  $g$ , a simple isomorphism can be set up between them which makes  $s'^{m-1}_0$  correspond to  $s''^{m-1}_0$ . The theorem in question then yields the following result. *If two  $m$ -groups reducible to cyclic polyadic groups are of the same order, and one  $m$ -group has an element of the same order as an element of the other, then the  $m$ -groups are simply isomorphic.*

Every automorphism of an  $m$ -group  $G$  permutes the elements of  $G$  according to a certain ordinary substitution. These substitutions clearly constitute an ordinary substitution group which may be called the group of isomorphisms of  $G$ . This terminology may be reconciled with that of §16 by noting that when  $G$  is represented as a regular substitution group, the corresponding  $(K_0)_{11}$  of §16 is simply isomorphic with the group of isomorphisms of  $G$ .

On the other hand the substitutions which result merely from transforming  $G$  by its own elements need not form a 2-group. In fact, it is readily verified that they do form an ordinary substitution group when and only when  $G$  has an invariant element. However they clearly do form an  $m$ -adic group of ordinary substitutions which may then be called the group of inner isomorphisms of  $G$ . It is easily proved that as in the ordinary theory this  $m$ -group is simply isomorphic with the central quotient group of  $G$ . Hence it is simply isomorphic with  $G$  if and only if the associated central of  $G$  is the identity.

By using the fact that every automorphism of  $G$  can be obtained by transforming it by some element, it is readily proved that the group of inner isomorphisms of  $G$  is an invariant subgroup of the group of isomorphisms of  $G$ , if not identical with it, when the latter is extended to an  $m$ -group. On the other hand, the containing group of the group of inner isomorphisms is directly an invariant subgroup of the group of isomorphisms, when not identical with it. This containing group clearly consists of the substitutions according to which the elements of  $G$  are permuted when  $G$  is transformed by all of its polyads.

In extending the Sylow subgroup property of the group of inner isomorphisms of an ordinary group to  $m$ -groups, we have to restrict our  $m$ -adic  $G$  to be of order  $g$  with  $g/p^a$  prime to  $m-1$ ,  $p^a$  being the largest power of the prime  $p$  dividing  $g$ . Since the order of  $I_{11}$ , the  $m$ -group of inner isomorphisms of  $G$ , divides  $g$ ,  $I_{11}$  has the same order property. We can then show that  $I_{11}$  contains the same number of Sylow subgroups corresponding to  $p$  as  $G$  does, it being understood that if  $p$  does not divide the order of  $I_{11}$ , the corresponding Sylow subgroups of  $I_{11}$  are its subgroups of first order. While the proof differs little from the corresponding ordinary group proof, we cannot follow Miller

in dismissing it with a line, and instead present it at least in outline. The elements of  $I_{11}$  corresponding to the elements of a subgroup  $H$  of  $G$  constitute a subgroup  $H'$  of  $I_{11}$  which may be called  $H$ 's corresponding subgroup. Let  $H$  be a Sylow subgroup of  $G$  for the prime  $p$  in accordance with our hypothesis. Then, by considering  $I_{11}$  to be the central quotient group of  $G$ , and comparing the largest powers of  $p$  dividing the orders of  $H$ ,  $I_{11}$ , and  $C_0$  with those dividing the orders of  $H$ ,  $H'$ , and the crosscut of  $H_0$  and  $C_0$ , we are enabled to conclude that  $H'$  is a Sylow subgroup of  $I_{11}$  for the prime  $p$ . Since corresponding elements of  $G$  and  $I_{11}$  transform corresponding subgroups into corresponding subgroups, the relation between the Sylow subgroups of  $G$  for the prime  $p$  and their corresponding subgroups of  $I_{11}$  is shown by the complete set of conjugates theorem to be a correspondence between all the Sylow subgroups of  $G$ , and all the Sylow subgroups of  $I_{11}$ , for the prime  $p$ . Finally, since any subgroup of  $G$  with given corresponding subgroup of  $I_{11}$  would be transformed into itself by any other subgroup of  $G$  with that corresponding subgroup of  $I_{11}$ , the above correspondence must be 1-1.

The fact that the central quotient group of a non-abelian group cannot be cyclic leads in ordinary group theory to the result that the order of the group of inner isomorphisms of a non-abelian group is at least four. In the case of a non-abelian  $m$ -group, the same theorem, used in conjunction with our determination of the  $m$ -groups of the first three orders, shows that the least order of the group of inner isomorphisms of  $m$ -groups is at least two when  $m-1$  is even, three when  $m-1$  is odd but divisible by 3, four when  $m-1$  is neither divisible by 2 nor 3. The following examples show that these actually are the least orders of  $I_{11}$  for such  $m$ 's as well as the fact that the order of  $I_{11}$  may have any value from that least order up to and including the order four. First, by extending an ordinary group with  $I_1$  of order four to an  $m$ -group, we see that for any  $m$ ,  $I_{11}$  may be of order four. An  $I_{11}$  of order three is immediately furnished for  $m-1$  even by the non-abelian  $m$ -group of order three itself. For  $m-1$  odd, but divisible by 3, we have the following example with  $m-1=3$ , and hence by extension for any  $m$  with  $m-1$  divisible by 3. Let  $G_0$  be the ordinary cyclic group of order nine generated by the cyclic substitution  $t = (a_1a_2a_3a_4a_5a_6a_7a_8a_9)$ . Then  $s = (a_2a_5a_8)(a_3a_6a_9)$  transforms  $t$  into  $t^4$  while  $s^3 = 1$ .  $G = G_0s$  is then a 4-group of order nine. Since  $G_0$  is abelian, the associated central  $C_0$  of  $G$  consists of the elements of  $G_0$  invariant under  $s$ , i.e., of 1,  $t^3$ ,  $t^6$ . The  $I_{11}$  of  $G$  is therefore also of order three. Finally an  $I_{11}$  of order two for  $m-1=2$ , and hence by extension for any even  $m-1$ , is exhibited by the following 3-group of order four. Let  $G_0$  be the axial group 1,  $(ab)$ ,  $(cd)$ ,  $(ab)(cd)$ ,  $s$  the substitution  $(ac)(bd)$ . Since  $s$  transforms  $G_0$  into itself, while  $s^2 = 1$ ,  $G = G_0s$  is a 3-group of order four. As  $s$  transforms but 1 and  $(ab)(cd)$  of  $G_0$  into themselves, the  $C_0$  of  $G$ , and hence also the  $I_{11}$  of  $G$ , is of order two.

When  $I_{11}$  is of order two it can abstractly be but the noncyclic  $m$ -group of

order two with its two first order elements.  $G$  is correspondingly separated into two abelian subgroups of half its order. It is readily proved that every abelian subgroup of  $G$  is contained in one of these subgroups. Conversely, if non-abelian  $G$  can be separated into two abelian subgroups, its  $I_{11}$  is of order two.

When  $I_{11}$  is of order three, it can be but the non-abelian group when  $m-1$  is of the form  $6\mu+2$  and  $6\mu+4$ , the abelian noncyclic group when  $m-1$  is of the form  $6\mu+3$ , and either of these two when  $m-1$  is of the form  $6\mu+6$  as shown by extensions of the cases where  $m-1=2$  and  $3$ . In any event  $I_{11}$  consists of three first order elements, so that  $G$  is separated into three abelian subgroups of one-third its order. Again every abelian subgroup of  $G$  is contained in one of these three subgroups. We have not however been able to decide the question whether a non-abelian  $G$  which can be separated into three abelian subgroups of one-third its order must have  $I_{11}$  of order three.

We restrict our discussion of  $I_{11}$  of order four to  $m$ 's for which four is the least order of  $I_{11}$ , i.e., to  $m-1$  not divisible by 2 or 3. Since  $m-1$  is then prime to the order of  $I_{11}$ , while the smallest prime divisor of  $m-1$  cannot be less than 5, our seemingly trivial form for the number of first order elements of an  $m$ -group with  $m-1$  prime to  $g$  shows that  $I_{11}$  has exactly one first order element.  $I_{11}$  is therefore reducible to an ordinary group of order four, and indeed to the axial group. Furthermore, the subgroups of  $I_{11}$  reduce to the subgroups of the axial group when  $I_{11}$  is so reduced. It follows that  $G$  then has three abelian subgroups of half its order, while every abelian subgroup of  $G$  is contained in one of these subgroups. Conversely, if a non-abelian  $m$ -group with  $m-1$  not divisible by 2 or 3 has more than one abelian subgroup of half its order, its  $I_{11}$  is reducible to the axial group.

**33. Extension of Frobenius's theorem to  $m$ -adic groups.** Thanks to recent work of Hall<sup>(29)</sup> on a wide generalization of Frobenius's theorem, the extension of the original theorem of Frobenius to polyadic groups is immediate. A very special case of Theorem III of Hall's paper may be stated as follows. If a subgroup  $H$  is transformed into itself by an element  $P$ , then the number of solutions of  $X^N=1$  which lie in the coset  $HP$  is congruent to 0 modulo H.C.F.  $(N, h)$ , where  $h$  is the order of  $H$ . Given, then, an arbitrary  $m$ -group  $G$  of order  $g$ , express  $G$  in the form  $G=G_0s_0$  in accordance with our coset theorem. With  $n$  a divisor of  $g$ , the elements  $s$  of  $G$  whose  $m$ -adic orders divide  $n$  are those for which  $s^{[n]}=s$ , i.e.,  $s^{(m-1)n}=1$ . Since  $G_0$  is transformed into itself under  $s_0$ , the above special case of Hall's theorem is immediately applicable to yield the following result. *The number of elements of an  $m$ -group  $G$  of order  $g$  whose ( $m$ -adic) orders divide an arbitrary divisor  $n$  of  $g$  is, if not 0, not only a multiple of  $n$ , but of  $n$  H.C.F.  $(g/n, m-1)$ .*

That the number in question may be 0 is shown by a cyclic  $m$ -group of

(<sup>29</sup>) P. Hall, *On a theorem of Frobenius*, Proceedings of the London Mathematical Society, (2), vol. 40 (1935-1936), pp. 468-501.

order  $g$  with  $g/n$  not prime to  $m-1$ . If  $\gamma$  is any divisor of  $n$ ,  $g/\gamma$  will also fail to be prime to  $m-1$ , and the cyclic group has no elements of orders dividing  $n$ . Note that when  $g$  is prime to  $m-1$  this can never occur, for our otherwise arbitrary  $G$  must then have at least one first order element. Actually, by applying the above result, restated for  $n$  not a divisor of  $g$ , to the conjugate subgroups of  $G$  of §26—and for these subgroups, indeed, the result is easily obtainable with but the help of the ordinary Frobenius theorem—we obtain the following stronger result. *If an  $m$ -group  $G$  is of order  $g$  prime to  $m-1$ , and  $n$  is any divisor of  $g$ , then the number of elements of  $G$  whose orders divide  $n$  is a multiple not only of  $n$ , but of  $n$  H.C.F.  $(g/n, \lambda)$ ,  $\lambda$  being the number of first order elements of  $G$ .*

**34. Representation of an abstract  $m$ -adic group as a transitive  $(m, \mu)$  substitution group.** We shall consider the general question of representing an abstract  $m$ -group  $G$  of order  $g$  by a transitive  $m$ -adic group of  $\mu$ -adic substitutions of degree  $n$ . (See §17.) The result can then immediately be specialized to the two cases of chief interest,  $\mu = m$  and  $\mu = 2$ , as well as to the case  $n = g$ , i.e., when the representing group is regular.

In the general case it is necessary to introduce polyadic groups intermediate between  $G$  and its associated ordinary group  $G_0$ , groups whose introduction simultaneously with that of  $G_0$  could have been used to generalize the theory at a number of points<sup>(90)</sup>. Clearly each coset in the expansion of the abstract containing group  $G^*$  of  $G$  as regards  $G_0$  is a polyadic group of order  $g$  under suitable extensions of the dyadic operation of  $G^*$ . In particular, if  $i$  is a divisor of  $m-1$ , the coset consisting of the  $i$ -ads of  $G$ , regarded as members of  $G^*$ , will thus constitute a group of dimension  $(m-1)/i+1$ . It will suffice to refer to this group as the polyadic group  $G_i$  of the  $i$ -ads of  $G$ . In particular  $G_1 = G$ ,  $G_{m-1} = G_0$ . As in the case of the subgroups of  $G$ , we may identify  $(G_i)^*$  with the subgroup of  $G^*$  generated by the elements of  $G_i$ .  $(G_i)_0$  is then simply  $G_0$ . Finally, since the isomorphism between  $G^*$  and any other containing group of  $G$  established in §6 involves but a 1-1 correspondence between the elements of two corresponding cosets,  $G_i$  may similarly be set up by means of any containing group of  $G$ .

Suppose then that  $G$  can be represented by a transitive  $(m, \mu)$  group  $G'$  of degree  $n$ , with, of course,  $\mu-1$  a divisor of  $m-1$ . Corresponding to the polyadic group  $G_{\mu-1}$  of the  $(\mu-1)$ -ads of  $G$  there will then be the polyadic group  $G'_{\mu-1}$  of the  $(\mu-1)$ -ads of  $G'$ , conveniently set up by means of the containing group of  $G'$  generated by the substitutions of  $G'$ .  $G'_{\mu-1}$  then consists of substitutions carrying each of the  $\mu-1$   $\Gamma$ 's on which  $G'$  is written into themselves<sup>(91)</sup>. Since  $G'$  is transitive, at least one substitution of  $G'_{\mu-1}$  carries  $a_{11}$

<sup>(90)</sup> E.g., see the end of the last footnote to §7. Likewise the concept of semi-invariant subgroups could correspondingly be generalized.

<sup>(91)</sup> Note that these will also be the substitutions forming  $G'_0$  when and only when the containing group generated by  $G'$  is of index  $\mu-1$ .



into itself. The set of all such substitutions in  $G'_{\mu-1}$  then constitutes a subgroup  $H'_{\mu-1}$  of  $G'_{\mu-1}$  of order  $g/n$ . The associated ordinary group  $H'_0$  of  $H'_{\mu-1}$  is a subgroup of the associated ordinary group  $G'_0$  of  $G'$ , and, in fact, consists of the substitutions of  $G'_0$  carrying  $a_{11}$  into itself. It then follows from the transitivity of  $G'$  that neither  $H'_0$ , if it be not the identity, nor any subgroup of  $H'_0$  other than the identity is invariant under  $G'$ .

It therefore follows that for  $G$  to be representable by a transitive  $(m, \mu)$  group of degree  $n$ ,  $\mu-1$  a divisor of  $m-1$ , it is necessary that  $G_{\mu-1}$  have a subgroup  $H_{\mu-1}$  of order  $g/n$  such that neither  $H_0$ , that is,  $(H_{\mu-1})_0$ , if it be not the identity, nor any subgroup of  $H_0$  other than the identity is invariant under  $G$ . We now prove this condition also sufficient. Each right coset of  $G^*$  as regards  $H_0$  consists of  $g/n$   $i$ -ads with fixed  $i$ .  $H_{\mu-1}^*$  consists of  $(m-1)/(\mu-1)$  of these cosets, one for each  $i$  a multiple of  $\mu-1$ . Each right coset of  $G^*$  as regards  $H_{\mu-1}^*$  therefore also consists of  $(m-1)/(\mu-1)$  of the right cosets of  $G^*$  as regards  $H_0$ , one for each  $i$  differing from a fixed  $i=i_0$  by a multiple of  $\mu-1$ . We may then choose  $i_0$  so that  $1 \leq i_0 \leq \mu-1$ . And for each such  $i_0$  there will be exactly  $n$  right cosets of  $G^*$  as regards  $H_{\mu-1}^*$  which together exhaust all  $i$ -ads with  $i-i_0$  a multiple of  $\mu-1$ . Now symbolize the  $n$  right cosets of  $G^*$  as regards  $H_{\mu-1}^*$  with  $i_0=1$  by the letters  $a_{11}, a_{12}, \dots, a_{1n}$ . These together will form the  $\Gamma_1$  of the basis of our representation. Similarly for  $\Gamma_2, \dots, \Gamma_{\mu-1}$ , with  $i_0$  correspondingly  $2, \dots, \mu-1$ . If now we multiply the elements of  $G^*$  on the right by an element  $s$  of  $G$ , the effect on the right cosets of  $G^*$  as regards  $H_{\mu-1}^*$  is merely to permute them as units, the  $i_0$  of such a coset becoming  $i_0+1$ , reduced modulo  $\mu-1$  if need be. In terms of the  $a$ 's therefore, the letters of  $\Gamma_1$  go over in 1-1 fashion into those of  $\Gamma_2$ , of  $\Gamma_2$  into those of  $\Gamma_3, \dots$ , of  $\Gamma_{\mu-1}$  into those of  $\Gamma_1$ . Corresponding to  $s$  there is thus determined a  $\mu$ -adic substitution  $s'$  of degree  $n$  on the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_{\mu-1}$ . The set of all such  $\mu$ -adic substitutions corresponding to elements of  $G$  clearly constitute an  $m$ -group  $G'$ , under the product of  $m$  substitutions as operation, isomorphic with  $G$ . This isomorphism is also simple. For if  $s_1$  and  $s_2$  are any two elements of  $G$  corresponding to the same substitution  $s'$  of  $G'$ ,  $t=s_1s_2^{-1}$  must be both in  $G_0$  and  $H_{\mu-1}^*$ , and hence in  $H_0$ . The set of such  $t$ 's must then be a group contained in  $H_0$ , and invariant under  $G$ , and hence consists of the identity only. That is,  $s_1=s_2$ .

We have thus proved the following theorem. *A necessary and sufficient condition that an abstract  $m$ -group  $G$  of order  $g$  can be represented as a transitive  $m$ -adic group of  $\mu$ -adic substitutions of degree  $n$ ,  $\mu-1$  a divisor of  $m-1$ , is that the polyadic group of  $(\mu-1)$ -ads of  $G$  contains a subgroup of order  $g/n$  whose associated ordinary group, if not the identity, is not invariant under  $G$ , and contains no subgroup besides the identity invariant under  $G$ . For the representation of  $G$  by a transitive  $m$ -adic substitution group of degree  $n$  this condition reduces to the condition that the associated ordinary group of  $G$  contains a subgroup of order  $g/n$  which, if not the identity, is not invariant under  $G$ , and*



$\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  with complex coefficients form an  $m$ -group under this operation. For the associative law follows immediately from this reinterpretation. Furthermore, if in the equation  $A_1 A_2 \dots A_m = A_{m+1}$  all but  $A_i$  are specified  $m$ -adic linear transformations,  $A_i$  will be determined as an ordinary linear transformation and be given by the equation  $A_i = A_{i-1}^{-1} \dots A_1^{-1} A_{m+1} A_m^{-1} \dots A_{i+1}^{-1}$ . Now each  $A^{-1}$  carries  $\Sigma_j$  into  $\Sigma_{j-1}$ . Hence  $A_i$  carries  $\Sigma_j$  into  $\Sigma_k$  where  $k \equiv j - (m-1) + 1 \pmod{m-1}$ , i.e.,  $k \equiv j + 1 \pmod{m-1}$ , and  $A_i$  is also an  $m$ -adic linear transformation.

We shall call any set of  $m$ -adic linear transformations of  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  which constitute an  $m$ -group under the above operation an  *$m$ -adic linear group in  $n$  variables*. Any such  $m$ -group will then be a subgroup of the above "complete"  $m$ -adic linear group in  $n$  variables. It follows that the necessary and sufficient condition that a finite set of  $m$ -adic linear transformations of  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  with complex coefficients form an  $m$ -adic linear group is that the product of any  $m$  members of the set is in the set. Unless otherwise indicated,  $m$ -adic linear group will mean finite  $m$ -adic linear group in the present paper. However, the infinite complete  $m$ -adic linear group is useful in serving as fundamental  $m$ -group for operations on arbitrary  $m$ -adic linear transformations. Its members, as ordinary linear transformations in  $(m-1)n$  variables, will generate a containing group of index  $m-1$  which may therefore be used in place of its abstract containing group. Its ordinary associated group, consisting of the products of  $m-1$   $m$ -adic linear transformations of  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ , will therefore consist of transformations which carry each  $\Sigma^{(i)}$  into itself, and indeed of all linear transformations with complex coefficients which carry each  $\Sigma^{(i)}$  into itself. We may therefore refer to such transformations as  $(m-1)$ -ads of  $m$ -adic linear transformations, or briefly  $(m-1)$ -ads.

While it will continue to be useful every so often to consider  $m$ -adic linear transformations as special forms of ordinary linear transformations, it is as generalization of ordinary linear transformation that they lend themselves to a corresponding generalization of the ordinary theory. For this purpose we return to our arbitrary  $m$ -adic linear transformation  $A$ , and as in ordinary theory represent it by the  *$m$ -adic matrix*

$$A = [A', A'', \dots, A^{(m-1)}],$$

where the component  $A^{(i)}$  is the ordinary matrix

$$A^{(i)} = \begin{pmatrix} a_{11}^{(i)} & a_{12}^{(i)} & \dots & a_{1n}^{(i)} \\ a_{21}^{(i)} & a_{22}^{(i)} & \dots & a_{2n}^{(i)} \\ \cdot & \cdot & \dots & \cdot \\ a_{n1}^{(i)} & a_{n2}^{(i)} & \dots & a_{nn}^{(i)} \end{pmatrix}$$



Clearly the identity among  $(m-1)$ -ads is  $(E, E, \dots, E)$ , where  $E$  is the ordinary matrix identity, while the inverse of  $(\alpha', \alpha'', \dots, \alpha^{(m-1)})$  is  $((\alpha')^{-1}, (\alpha'')^{-1}, \dots, (\alpha^{(m-1)})^{-1})$ .

We consider now the important question of change of variable. Let  $S$  be an  $m$ -adic linear transformation carrying the  $x_{ij}$ 's into the  $x_{(i+1)k}$ 's,  $T$  an  $m$ -adic linear transformation expressing the  $x_{ij}$ 's in terms of  $X_{(i+1)k}$ 's, and likewise the  $x_j$ 's in terms of  $X'_{(i+1)k}$ 's. As a result, the  $X_{ij}$ 's are carried into the  $X'_{(i+1)k}$ 's according to an  $m$ -adic linear transformation  $R$ . We shall say that  $R$  is the result of  $m$ -adically changing variables in  $S$  according to  $T$ . Now with  $R$ ,  $S$ , and  $T$  considered to be ordinary linear transformations on  $(m-1)n$  variables,  $R$  is the result of an ordinary change of variables in  $S$  according to  $T$ , and hence is the transform of  $S$  with respect to  $T$ . If then in the equation  $R = T^{-1}ST$  we follow through the successive linear transformations, we obtain the following results on the corresponding  $m$ -adic matrices. If

$$S = [S', S'', \dots, S^{(m-1)}], \quad T = [T', T'', \dots, T^{(m-1)}],$$

then the transform

$$R = [R', R'', \dots, R^{(m-1)}]$$

of  $S$  with respect to  $T$ , which is the result of  $m$ -adically changing the variables of  $S$  according to  $T$ , is given by the equations

$$R^{(i)} = [T^{(i-1)}]^{-1} S^{(i-1)} T^{(i)}, \quad i = 1, 2, \dots, m-1 \quad (98).$$

Closer to the ordinary concept of change of variable would be instituting an ordinary change of variable in each space  $\Sigma^{(i)}$ . This would then correspond to changing variables according to an  $(m-1)$ -ad. As before, if  $S$  is an  $m$ -adic linear transformation,  $\tau$  equivalent to an  $(m-1)$ -ad of  $m$ -adic linear transformations, the result of changing variables in  $S$  according to  $\tau$  will be an  $m$ -adic linear transformation  $R$  with  $R = \tau^{-1}S\tau$ . The corresponding formula for transforming the  $m$ -adic matrix  $S = [S', S'', \dots, S^{(m-1)}]$ , by the  $(m-1)$ -ad  $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$  to yield the  $m$ -adic matrix  $R = [R', R'', \dots, R^{(m-1)}]$  may again be obtained by following through the transformations involved, or, perhaps just as easily, by applying our formulas for operations on  $(m-1)$ -ads. We thus obtain

$$R^{(i)} = [\tau^{(i)}]^{-1} S^{(i)} \tau^{(i+1)}, \quad i = 1, 2, \dots, m-1.$$

While our  $m$ -adic matrix notation is more convenient in most applications, our later generalization of characteristic equation requires rather the matrix of the corresponding ordinary linear transformation in the  $(m-1)n$  variables.

(98) These equations can also be obtained from the equations defining the  $m$ -adic operation on  $m$ -adic matrices, and the original  $m$ -adic definition of transform.



With  $A = [A', A'', \dots, A^{(m-1)}]$ , the corresponding ordinary matrix then has the following form

$$\begin{pmatrix} 0 & A' & 0 & \dots & 0 \\ 0 & 0 & A'' & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & A^{(m-2)} \\ A^{(m-1)} & 0 & 0 & \dots & 0 \end{pmatrix}.$$

If then  $D$  is the determinant of this matrix,  $D', \dots, D^{(m-1)}$  of the components  $A', \dots, A^{(m-1)}$  of  $A$ , it follows that

$$D = (-1)^{mn} D' D'' \dots D^{(m-1)}.$$

By contrast, for the  $(m-1)$ -ad  $\alpha = (\alpha', \alpha'', \dots, \alpha^{(m-1)})$ , the corresponding ordinary matrix has the components of  $\alpha$  along its principal diagonal, zero's elsewhere, and the determinant of the matrix is always the product of the determinants of the components.

**36.  $m$ -adic collineations and collineation-groups.** If the variables of each space  $\Sigma^{(i)}$  be considered homogeneous coordinates in a corresponding space  $S^{(i)}$  of dimension  $n-1$ , our  $m$ -adic linear transformation  $A$  on  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ , may be said to define an  $m$ -adic collineation on  $S', S'', \dots, S^{(m-1)}$ . In fact, if we let the ratios  $x_{i1}/x_{in}, \dots, x_{i(n-1)}/x_{in}$  be denoted by  $y_{i1}, \dots, y_{i(n-1)}$ , we are thus led to the  $m$ -adic linear fractional transformation  $i=1, 2, \dots, m-1$ :

$$y_{is} = \frac{a_{s1}^{(i)} y'_{(i+1)1} + \dots + a_{s(n-1)}^{(i)} y'_{(i+1)(n-1)} + a_{sn}^{(i)}}{a_{n1}^{(i)} y'_{(i+1)1} + \dots + a_{nn}^{(i)} y'_{(i+1)(n-1)} + a_{nn}^{(i)}}, \quad s = 1, 2, \dots, n-1.$$

Unlike the case of an  $m$ -adic linear transformation, our  $m$ -adic linear fractional transformation is in general not a special case of an ordinary linear fractional transformation on all the variables, since the denominators in general are not all the same. On the other hand it justifies our phrase  $m$ -adic collineation, since the equality of the denominators for each  $i$  insures our  $m$ -adic linear fractional transformation on the nonhomogeneous  $y$ 's carrying the straight lines of each  $S^{(i)}$  into those of  $S^{(i+1)}$ . Moreover, the product of  $m$   $m$ -adic linear fractional transformations of  $S', S'', \dots, S^{(m-1)}$  will again be of that form, so that we can expect to have  $m$ -adic linear fractional groups, and hence  $m$ -adic collineation-groups.

Two  $m$ -adic linear transformations  $A_1$  and  $A_2$  on  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  will yield the same  $m$ -adic linear fractional transformation on  $S', S'', \dots, S^{(m-1)}$  when and only when their  $m$ -adic matrices  $A_1 = [A'_1, A''_1, \dots, A_1^{(m-1)}]$  and  $A_2 = [A'_2, A''_2, \dots, A_2^{(m-1)}]$  are such that the elements of each component  $A_1^{(i)}$  are a constant  $k_i$  times the elements of the corresponding component  $A_2^{(i)}$ .

This then is the condition that  $A_1$  and  $A_2$  represent the same  $m$ -adic collineation. Since the  $k_i$ 's need not be the same,  $A_1$  and  $A_2$  as ordinary linear transformations need not then represent the same collineation in the ordinary sense. If now we let  $\tau$  be the  $(m-1)$ -ad

$$((k_1, k_1, \dots, k_1), (k_2, k_2, \dots, k_2), \dots, (k_{m-1}, k_{m-1}, \dots, k_{m-1}))$$

whose components are all ordinary similarity-matrices, we see from the preceding section that  $A_1 = \tau A_2$ . We shall call an  $(m-1)$ -ad each of whose components is a similarity-matrix a *similarity- $(m-1)$ -ad*. It follows that  $A_1$  and  $A_2$  represent the same  $m$ -adic collineation when and only when  $A_1 A_2^{-1}$  is a similarity- $(m-1)$ -ad.

$A_2^{-1} A_1$  must then also be a similarity- $(m-1)$ -ad; but it will equal  $A_1 A_2^{-1}$  when and only when the  $k_i$ 's are all equal. In fact, again by the preceding section, writing the above  $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$ , we find that  $A_1 = A_2 \bar{\tau}$ , where  $\bar{\tau} = (\tau^{(m-1)}, \tau', \dots, \tau^{(m-2)})$ , and hence  $A_2^{-1} A_1 = \bar{\tau}$ . Comparing these two results, we see that  $A_2^{-1} \tau A_2 = \bar{\tau}$ . Since  $A_2$  is an arbitrary  $m$ -adic matrix, it follows that every  $m$ -adic matrix transforms a similarity- $(m-1)$ -ad  $(\tau', \tau'', \dots, \tau^{(m-1)})$  into the similarity- $(m-1)$ -ad

$$(\tau^{(m-1)}, \tau', \dots, \tau^{(m-2)}).$$

By contrast, every similarity- $(m-1)$ -ad is transformed into itself by an  $(m-1)$ -ad.

Consider now any  $m$ -adic linear group  $G$ . Since the product of two similarity- $(m-1)$ -ads is again a similarity- $(m-1)$ -ad, the similarity- $(m-1)$ -ads of  $G_0$ , the associated ordinary group of  $G$ , will constitute a subgroup  $H_0$  of  $G_0$ . Since every  $m$ -adic matrix transforms a similarity- $(m-1)$ -ad into a similarity- $(m-1)$ -ad,  $H_0$  will be invariant under  $G$ . We may therefore form the  $m$ -adic quotient group  $K = G/H_0$ . Each coset of  $G$  as regards  $H_0$  can be written  $H_0 A$  with  $A$  in  $G$ , and hence consists of elements of  $G$  representing the same  $m$ -adic collineation as  $A$ , and, in fact, of all such elements of  $G$ . The elements of  $K$  are thus in 1-1 correspondence with the distinct  $m$ -adic collineations represented by the elements of  $G$ .  $K$  may therefore be called the  *$m$ -adic collineation-group* corresponding to  $G$ .

An arbitrary  $m$ -adic collineation-group  $G$  may be given by arbitrarily representing each collineation by an  $m$ -adic linear transformation<sup>(24)</sup>. If  $G$  is of order  $g$ , and written thus "on  $n$  variables," a modification of the ordinary treatment will yield an  $m$ -adic linear group of order  $n^{m-1}g$  which is  $(n^{m-1}, 1)$  isomorphic with  $G$ , and whose transformations have *components of determinant unity*. In fact let  $S = [S', S'', \dots, S^{(m-1)}]$  be in  $G$  thus represented, with the determinants of its components  $D', D'', \dots, D^{(m-1)}$  respectively. Let  $\theta^{(i)}$  be any solution of the equation  $[\theta^{(i)}]^n = [D^{(i)}]^{-1}$ , and form the similarity-

<sup>(24)</sup> The product of  $m$  such representatives need not then be in the given set of representatives, but need merely represent the same  $m$ -adic collineation as some member of the set.

$(m-1)$ -ad  $\tau = ((\theta', \theta', \dots, \theta'), (\theta'', \theta'', \dots, \theta''), \dots, (\theta^{(m-1)}, \theta^{(m-1)}, \dots, \theta^{(m-1)}))$ . Then  $A = \tau S = [(\theta', \theta', \dots, \theta')S', (\theta'', \theta'', \dots, \theta'')S'', \dots, (\theta^{(m-1)}, \theta^{(m-1)}, \dots, \theta^{(m-1)})S^{(m-1)}]$  represents the same  $m$ -adic collineation as  $S$ , and has all of its components of determinant unity. For each  $S$  there will thus be  $n^{m-1}$   $A$ 's, and these constitute all of the  $m$ -adic linear transformations with components of determinant unity representing the same  $m$ -adic collineation as  $S$ . It then readily follows that the set of  $n^{m-1}g$   $m$ -adic linear transformations thus corresponding to the  $g$  elements of  $G$  constitute a linear  $m$ -group isomorphic with  $G$ . For let  $S_1, S_2, \dots, S_m$  be any  $m$  transformations in the original representation of  $G$ ,  $A_1 = \tau_1 S_1, A_2 = \tau_2 S_2, \dots, A_m = \tau_m S_m$  corresponding transformations with components of determinant unity. Then

$$A = A_1 A_2 \cdots A_m$$

has for its  $i$ th component

$$A^{(i)} = A_1^{(i)} A_2^{(i+1)} \cdots A_m^{(i)} = \tau_1^{(i)} \tau_2^{(i+1)} \cdots \tau_m^{(i)} S_1^{(i)} S_2^{(i+1)} \cdots S_m^{(i)} = \tau^{(i)} S^{(i)},$$

where  $S = S_1 S_2 \cdots S_m$ , and  $\tau$  is a similarity- $(m-1)$ -ad.  $A$  therefore has components of determinant unity, and represents the same  $m$ -adic collineation as  $S$ .  $A$  is therefore in our set of  $n^{m-1}g$  transformations, whence finally our result.

To compare the ordinary treatment with this modification of it, we introduce the following considerations. Given an  $m$ -adic linear group  $G$ , those similarity- $(m-1)$ -ads of  $G_0$  which have equal components themselves constitute a subgroup  $H'_0$  of  $G_0$  invariant under  $G$ . We may therefore form the  $m$ -adic quotient group  $K' = G/H'_0$ . Each coset of the expansion of  $G$  as regards  $H'_0$  consists of all transformations in  $G$  which as ordinary transformations on  $(m-1)n$  variables correspond to the same ordinary collineation. We shall therefore call  $K'$  the collineation- $m$ -adic group of  $G$ . If now an arbitrary collineation- $m$ -adic group  $G$  be given by corresponding representative  $m$ -adic linear transformations, the ordinary treatment applies without modification; and if  $G$  is of order  $g$ , and on  $n$  variables, a linear  $m$ -adic group of order  $(m-1)ng$  is thus obtained which is  $[(m-1)n, 1]$  isomorphic with  $G$ , and whose members as ordinary transformations are of determinant unity. On the other hand, if an arbitrary  $m$ -adic collineation-group  $G$  be thus given, the ordinary unmodified treatment will in general be inapplicable. In fact, otherwise, the given representatives of the members of  $G$  must also be representatives of the members of a collineation- $m$ -adic group. This will clearly not be so for random representations of the members of  $G$ . And the following example shows that the  $m$ -adic collineation-group  $G$  may be such that no representation thereof will represent a collineation- $m$ -adic group. The triadic collineations corresponding to

$$A: \quad [(1, 1), (1, -1)], \quad B: \quad \left[ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right]$$

generate a triadic collineation-group  $G$  of order 4. The most arbitrary representations of  $A$  and  $B$  are

$$A: [(a, a), (b, -b)], \quad B: \left[ \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix}, \begin{pmatrix} 0 & d \\ -d & 0 \end{pmatrix} \right].$$

By direct computation we find that  $AAA = [(a^2b, -a^2b), (ab^2, ab^2)]$ ,  $BBA = [(-acd, acd), (bcd, bcd)]$ . As triadic collineations,  $AAA$  and  $BBA$  are identical, being the same as  $[(1, -1), (1, 1)]$ . As ordinary collineations, they can but be identified with  $[(a, -a), (b, b)]$ ,  $[(-a, a), (b, b)]$  which are never the same. Since any representation of  $G$  can have but one triadic linear transformation for each triadic collineation in  $G$ , no representation of this triadic collineation-group can also represent a collineation-triadic group.

If however  $G$  itself is an  $m$ -adic linear group, both methods are applicable. The unmodified treatment will then yield an  $m$ -adic linear group which is  $[(m-1)n, 1]$  isomorphic with the collineation- $m$ -adic group of  $G$ , and whose members as ordinary linear transformations have determinants unity. On the other hand, our modified treatment yields an  $m$ -adic linear group which is  $(n^{m-1}, 1)$  isomorphic with the  $m$ -adic collineation-group of  $G$ , and whose members have components of determinant unity.

**37.  $m$ -adic Hermitian invariants.** A set of  $m-1$  positive-definite Hermitian forms  $J = [J', J'', \dots, J^{(m-1)}]$ , one for each space  $\Sigma^{(i)}$ , will be said to be an  $m$ -adic (positive-definite) Hermitian form. Now

$$J^{(i)} = \sum_{k=1}^n \sum_{l=1}^n q_{kl}^{(i)} x_{ik} \bar{x}_{il}, \quad q_{lk}^{(i)} = \bar{q}_{kl}^{(i)},$$

can be transformed into

$$I^{(i)} = y_{i1} \bar{y}_{i1} + y_{i2} \bar{y}_{i2} + \dots + y_{in} \bar{y}_{in}$$

by a change of variables of the form

$$y_{ik} = \sum_{l=1}^k \rho_{kl}^{(i)} x_{il}, \quad k = 1, 2, \dots, n.$$

Hence  $J = [J', J'', \dots, J^{(m-1)}]$  can be transformed into  $I = [I', I'', \dots, I^{(m-1)}]$  by changing variables in  $\Sigma'$ ,  $\Sigma''$ ,  $\dots$ ,  $\Sigma^{(m-1)}$  according to an  $(m-1)$ -ad whose components, with  $i = 1, 2, \dots, m-1$ , are of the above form. The  $(m-1)$ -ad, of course, is that obtained by solving for the  $x$ 's in terms of the  $y$ 's. It is further understood that in operating on  $J$  by this  $(m-1)$ -ad, if  $x_{ij}$  is replaced by a certain expression,  $\bar{x}_{ij}$  is replaced by the conjugate of that expression.

If, on the other hand,  $J$  is transformed according to an  $m$ -adic change of variables,  $J^{(i)}$ , written on the variables of  $\Sigma^{(i)}$ , becomes an expression in the new variables not of  $\Sigma^{(i)}$  but of  $\Sigma^{(i+1)}$ . We are thus led to define an  $m$ -adic Hermitian invariant of an  $m$ -adic linear group as an  $m$ -adic Hermitian form

$J = [J', J'', \dots, J^{(m-1)}]$  such that each transformation in the group carries  $J' \rightarrow J'', J'' \rightarrow J''', \dots, J^{(m-1)} \rightarrow J'$ . It then readily follows that every  $m$ -adic linear group  $G$  has an  $m$ -adic Hermitian invariant. For let  $G'_0$  be the  $\Sigma'$  constituent group of  $G_0$ , the complete analogue of the  $G'_0$  of an  $m$ -adic substitution group. The linear group  $G'_0$  then has an Hermitian invariant  $J'$  on the variables of  $\Sigma'$ . Let  $S$  be in  $G$ , and let  $J''$  be the result of transforming  $J'$  according to  $S, \dots, J^{(m-1)}$  of transforming  $J^{(m-2)}$  according to  $S$ . Then  $J = [J', J'', \dots, J^{(m-1)}]$  will be an  $m$ -adic Hermitian form on  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ , and, as in §39 to come, is seen to be an  $m$ -adic Hermitian invariant of  $G$ .

By combining the above two results it follows that the variables of an  $m$ -adic linear group  $G$  may be so changed according to an  $(m-1)$ -ad that  $I = [I', I'', \dots, I^{(m-1)}], I^{(i)} = x_{i1}x_{i1} + x_{i2}x_{i2} + \dots + x_{in}x_{in}$ , is an  $m$ -adic Hermitian invariant of the resulting transform of  $G$ .

An  $m$ -adic linear group  $G$  in  $n$  variables will be said to be linearly reducible<sup>(\*)</sup> if by a suitable change of variables according to an  $(m-1)$ -ad there will be in  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  subspaces<sup>(\*\*)</sup>  $\Sigma'_1, \Sigma''_1, \dots, \Sigma^{(m-1)}_1$  respectively on  $\nu < n$  variables each such that  $\Sigma'_1 \rightarrow \Sigma''_1 \rightarrow \dots \rightarrow \Sigma^{(m-1)}_1 \rightarrow \Sigma'_1$  under every transformation in the resulting transform of  $G$ . If for some such change of variables the subspaces  $\Sigma'_2, \Sigma''_2, \dots, \Sigma^{(m-1)}_2$  on the remaining  $n - \nu$  variables of  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  are also each transformed into the next, then  $G$  will be said to be intransitive. In the first case  $\Sigma'_1, \Sigma''_1, \dots, \Sigma^{(m-1)}_1$  will be said to be a reduced set for  $G$ , in the second case a set of intransitivity of  $G$ . We then prove the theorem a linearly reducible  $m$ -adic linear group  $G$  is intransitive, and a reduced set constitutes one of the sets of intransitivity of  $G$ , subject, of course, to a change of variables in the reduced set according to an  $(m-1)$ -ad thereon. We may assume the variables in the reduced set to be the first  $\nu$  variables of each  $\Sigma^{(i)}$ . Then  $G$  may be further transformed by an  $(m-1)$ -ad so that it will have the  $m$ -adic Hermitian invariant  $I$  above. And this further change of variables, according to the form given above, merely transforms the reduced set according to an  $(m-1)$ -ad on its variables. With  $G$  in this last form, consider its containing group  $G^*$ . Then  $I^* = I' + I'' + \dots + I^{(m-1)}$  will be an ordinary Hermitian invariant of the ordinary linear group  $G^*$ , while the  $(m-1)\nu$  variables constituting the reduced set for  $G$  form a reduced set for  $G^*$  without further transformation. But then  $G^*$  is in intransitive form with those  $(m-1)\nu$  variables constituting a set of intransitivity of  $G^*$ . The same is then true of  $G$ .

An  $m$ -adic matrix  $A = [A', A'', \dots, A^{(m-1)}]$  will be said to be in canonical form if each component  $A^{(i)}$  is in the canonical form  $(a^{(i)}_1, a^{(i)}_2, \dots, a^{(i)}_n)$ . Then the corresponding ordinary theorem generalizes, i.e., if  $A$  is of finite

(\*) To distinguish between this extension of the ordinary concept and the totally unrelated polyadic concept we have termed reducibility.

(\*\*) Strictly, a misnomer, but a convenient one.



*m*-adic order, then it can be reduced to canonical form by transformation by an  $(m-1)$ -ad<sup>(97)</sup>. We shall prove this result in the next section more expeditiously. However we here give the analogue of the ordinary proof for the sake of the concepts thus introduced.

We prove then that we can always find  $m-1$  linear functions

$$y_{i1} = b_1^{(i)} x_{i1} + b_2^{(i)} x_{i2} + \cdots + b_n^{(i)} x_{in}, \quad i = 1, 2, \dots, m-1,$$

such that each  $y_{i1}$  is transformed into a constant  $\theta_i$  times  $y_{(i+1)1}$  by  $A$ . These  $m-1$  functions may then be said to constitute a *relative m-adic invariant* of  $A$ . With  $A$  the transformation

$$x_{is} = \sum_{t=1}^n a_{st}^{(i)} x'_{(i+1)t}, \quad s = 1, 2, \dots, n; \quad i = 1, 2, \dots, m-1,$$

we find that  $(y_{i1})A = \theta_i y_{(i+1)1}$  provided the following equations are true:

$$\theta_i b_t^{(i+1)} = \sum_{s=1}^n b_s^{(i)} a_{st}^{(i)}, \quad t = 1, 2, \dots, n.$$

By successive substitution, with  $i=1, 2, \dots, m-1$ , we obtain from these equations

$$\theta_1 \theta_2 \cdots \theta_{m-1} b_t' = \sum_{s=1}^n b_s' a_{st}^{(0)}, \quad t = 1, 2, \dots, n,$$

where the ordinary matrix  $(a_{st}^{(0)}) = A_0 = A'A'' \cdots A^{(m-1)}$ <sup>(98)</sup>. A set of solutions  $b_1', b_2', \dots, b_n'$ , not all zero, of this last set of equations can always be found provided  $\theta_1 \theta_2 \cdots \theta_{m-1}$  is a root of the characteristic equation of  $A_0$ . The preceding equations, with  $i=1, 2, \dots, m-2$ , then determine the remaining  $b$ 's, while the equations for  $i=m-1$  are then automatically satisfied.

Having thus found a relative *m*-adic invariant of  $A$ , the remainder of the proof follows the lines of the standard proof. That is, by a change of variables according to an  $(m-1)$ -ad given in part by our relative *m*-adic invariant of  $A$ , the new variables  $y_{11}, y_{21}, \dots, y_{(m-1)1}$  are transformed according to the equations  $y_{i1} = \theta_i y_{(i+1)1}$ ,  $i=1, 2, \dots, m-1$ , and hence constitute a reduced set for the *m*-adic linear group generated by  $A$ . If then  $A$  is of finite *m*-adic order, further change of variables according to an  $(m-1)$ -ad will

<sup>(97)</sup> It might be thought that since  $A$  as ordinary linear transformation is then of finite ordinary order, the standard theorem would apply. But note that an *m*-adic matrix in canonical form is not in canonical form as ordinary matrix. And from the contrary point of view, while  $A$  as ordinary matrix could thus be reduced to ordinary canonical form, the resulting linear transformation would no longer be an *m*-adic linear transformation; and the transformation used to obtain it would be a linear transformation on all the  $(m-1)n$  variables in a form constituting a meaningless jumble from the point of view of *m*-adic linear transformations.

<sup>(98)</sup> Or, more expeditiously, from  $(y_{11})A^{m-1} = \theta_1 \theta_2 \cdots \theta_{m-1} y_{11}$ .

change  $y_{11}, y_{21}, \dots, y_{(m-1)1}$  into a set of intransitivity of the group generated by  $A$ .  $A$  then determines an  $m$ -adic linear transformation on the remaining  $n-1$  variables, and the process may be repeated until  $A$  appears in canonical form, and, indeed, as the result of a single change of its original variables according to an  $(m-1)$ -ad.

Our proof of the existence of relative  $m$ -adic invariants of  $A$  might have taken a different turn. Our original  $(m-1)n$  homogeneous linear equations in the  $(m-1)n$  undetermined  $b$ 's will have a set of solutions not all zero, and hence, as shown by the equations themselves, not all zero for any  $i$ , provided the determinant of their coefficients is zero. We are thus led to one equation in the  $m-1$  unknowns  $\theta_1, \theta_2, \dots, \theta_{m-1}$  which may be called the  $m$ -adic characteristic equation of  $A$ . Its right-hand member is zero; left, the determinant of  $A$  as ordinary linear transformation with the elements of the principal diagonal, all zero in  $A$ , replaced by  $-\theta_{m-1}, \dots, -\theta_{m-1}, -\theta_1, \dots, -\theta_1, \dots, -\theta_{m-2}, \dots, -\theta_{m-2}$ . With  $\theta_1 = \theta_2 = \dots = \theta_{m-1} = \theta$ , the  $m$ -adic characteristic equation of  $A$  becomes the ordinary characteristic equation of  $A$  as ordinary linear transformation. We are thus, in fact, assured of relative  $m$ -adic invariants of  $A$  with  $\theta$ 's all equal. However, comparison with the earlier treatment yields the following result. The solutions of the  $m$ -adic characteristic equation of  $A = [A', A'', \dots, A^{(m-1)}]$  consist of all sets of values  $\theta_1, \theta_2, \dots, \theta_{m-1}$  for which  $\theta_1 \theta_2 \dots \theta_{m-1}$  is a root of the characteristic equation of  $A_0 = A' A'' \dots A^{(m-1)}$ .

**38. Reduction to canonical form.** If for two  $m$ -adic linear transformations  $A$  and  $B$  in  $n$  variables there is a third  $C$  such that  $B = C^{-1}AC$ , then  $A$  and  $B$  will be said to be *conjugate*. This is equivalent to there being an  $(m-1)$ -ad  $\gamma$  such that  $B = \gamma^{-1}A\gamma$ , since  $C$  and  $A^{m-2}C$  on the one hand,  $\gamma$  and  $A\gamma$  on the other, yield the same transform of  $A$ . It follows that the relation " $A$  and  $B$  are conjugate" is an equivalence relation. Likewise for  $m$ -adic linear groups.

The following easily proved theorem reduces the problem of conjugate  $m$ -adic linear transformations in  $n$  variables to that of conjugate ordinary linear transformations in  $n$  variables. *The necessary and sufficient condition that  $A = [A', A'', \dots, A^{(m-1)}]$  and  $B = [B', B'', \dots, B^{(m-1)}]$  are conjugate is that  $A_0 = A' A'' \dots A^{(m-1)}$  and  $B_0 = B' B'' \dots B^{(m-1)}$  are conjugate.* In fact, if  $B = \gamma^{-1}A\gamma$ ,  $\gamma = (\gamma', \gamma'', \dots, \gamma^{(m-1)})$ , then by our formula for change of variables according to an  $(m-1)$ -ad

$$B' = [\gamma']^{-1}A'\gamma'', B'' = [\gamma'']^{-1}A''\gamma''', \dots, B^{(m-1)} = [\gamma^{(m-1)}]^{-1}A^{(m-1)}\gamma'.$$

Hence

$$B'B'' \dots B^{(m-1)} = [\gamma']^{-1}A'A'' \dots A^{(m-1)}\gamma',$$

whence the necessity of our condition. Conversely, if  $A_0$  and  $B_0$  are conjugate,  $\gamma'$  may be chosen to satisfy the last of the above equations. If then  $\gamma'', \gamma''', \dots, \gamma^{(m-1)}$  are determined in accordance with the first  $m-2$  of the

change of variable equations, the last of those equations will be automatically satisfied. An  $(m-1)$ -ad  $\gamma = (\gamma', \gamma'', \dots, \gamma^{(m-1)})$  is thus determined which transforms  $A$  into  $B$ .

This result contrasts strongly with the corresponding result for  $(m-1)$ -ads. We may define two  $(m-1)$ -ads  $\alpha$  and  $\beta$  to be conjugate if there is an  $(m-1)$ -ad  $\gamma$  such that  $\beta = \gamma^{-1}\alpha\gamma$ . From our formula for the product of two  $(m-1)$ -ads it follows that  $\alpha = (\alpha', \alpha'', \dots, \alpha^{(m-1)})$  and  $\beta = (\beta', \beta'', \dots, \beta^{(m-1)})$  are conjugate when and only when the corresponding components  $\alpha^{(i)}$  and  $\beta^{(i)}$  are conjugate for each  $i$ . Hence, while the question of conjugacy for an  $m$ -adic matrix in  $n$  variables depends on but one ordinary matrix in  $n$  variables, the same question for an  $(m-1)$ -ad depends on  $m-1$  independent ordinary matrices in  $n$  variables each. Intrinsically, therefore, an  $m$ -adic matrix is far simpler than an  $(m-1)$ -ad. This is rather surprising in that apart from change of variables they are of equal generality; for if  $A$  is a fixed  $m$ -adic matrix the relation  $S = \tau A$  gives a 1-1 correspondence between all  $m$ -adic matrices  $S$  and  $(m-1)$ -ads  $\tau$ .

A more symmetrical though less useful condition for the  $m$ -adic matrices  $A$  and  $B$  being conjugate is that the  $(m-1)$ -ads  $A^{m-1}$  and  $B^{m-1}$  are conjugate. In fact, if  $A^{m-1} = \alpha$ , the equation  $A^m = \alpha A$  yields

$$A^{m-1} = (A'A'' \dots A^{(m-1)}, A''A''' \dots A', \dots, A^{(m-1)}A' \dots A^{(m-2)}).$$

The first component of  $A^{m-1}$  is therefore the  $A_0$  of our previous condition, while all the components are conjugate. The present condition then follows. We may note that all the components of an  $(m-1)$ -ad being conjugate is sufficient as well as necessary for the  $(m-1)$ -ad being the  $(m-1)$ -st ordinary power of some  $m$ -adic matrix. Intrinsically, then, an  $m$ -adic matrix is of the same degree of generality as an  $(m-1)$ -ad with conjugate components. Too much emphasis, however, must not be placed on the forms assumed by a single element under transformation, our present concern.

Returning to our first condition for the conjugacy of  $m$ -adic matrices, we have immediately that  $A = [A', A'', \dots, A^{(m-1)}]$  is conjugate to  $[A_0, E, \dots, E]$ , with  $A_0 = A'A'' \dots A^{(m-1)}$ . If now  $A$  is of finite  $m$ -adic order, then  $A^{m-1}$ , and hence its first component  $A_0$ , is of finite order.  $A_0$  is then conjugate to a matrix in the canonical form  $(a_1, a_2, \dots, a_n)$ . Hence, if  $A$  is of finite  $m$ -adic order, it is conjugate to an  $m$ -adic matrix in the canonical form  $[(a_1, a_2, \dots, a_n), E, \dots, E]$ .

More generally, if  $A$  is of finite  $m$ -adic order, it is conjugate to those  $m$ -adic matrices in the canonical form  $[(a_1', a_2', \dots, a_n'), (a_1'', a_2'', \dots, a_n''), \dots, (a_1^{(m-1)}, a_2^{(m-1)}, \dots, a_n^{(m-1)})]$  for which  $a_i' a_i'' \dots a_i^{(m-1)} = a_{j_1} a_{j_2} \dots a_{j_n}$  a permutation of  $a_1, a_2, \dots, a_n$ . Since  $a_1, a_2, \dots, a_n$  are the roots of the characteristic equation of  $A_0$ , we may say, as a consequence of the last section, that an  $m$ -adic matrix  $A$  of finite order assumes those canonical forms for which each selection of corresponding elements chosen from its components

constitutes a solution of the  $m$ -adic characteristic equation of  $A$ , while the corresponding roots of the characteristic equation of  $A_0$  are all of its roots each with the correct multiplicity. In particular, we may make  $a'_i = a''_i = \dots = a_i^{(m-1)}$  for each  $i$ . Hence the useful special result if  $A$  is of finite  $m$ -adic order, it is conjugate to an  $m$ -adic matrix in canonical form having equal components.

The most satisfactory generalization of an ordinary similarity-matrix is our similarity- $(m-1)$ -ad. An  $m$ -adic matrix each of whose components is a similarity-matrix will not in general remain of that form under transformation by an  $m$ -adic matrix<sup>(9)</sup>. We therefore define an  $m$ -adic similarity-matrix as one which is conjugate to an  $m$ -adic matrix whose components are all similarity-matrices. It readily follows from our criterion for the conjugacy of  $m$ -adic matrices that  $A = [A', A'', \dots, A^{(m-1)}]$  is an  $m$ -adic similarity-matrix when and only when  $A'A'' \dots A^{(m-1)}$  is a similarity-matrix. In particular, every first order  $m$ -adic matrix is an  $m$ -adic similarity-matrix. In fact,  $A$  is of  $m$ -adic order one when and only when  $A'A'' \dots A^{(m-1)} = E$ . Hence the first order  $m$ -adic matrices are the conjugates of  $[E, E, \dots, E]$ .

Our chief reason for introducing the above concept is the following theorem. If an  $m$ -adic linear group has an  $m$ -adic similarity-matrix as invariant element, it is conjugate to a group in which each element is an  $m$ -adic matrix with equal components. By an  $m$ -adic change of variable the invariant similarity-matrix can be transformed into an  $m$ -adic matrix  $A$  in canonical form in which the components are now equal similarity-matrices. If the given group is correspondingly transformed, a conjugate group having  $A$  as invariant element is obtained. For each element  $B$  of the transformed group we thus have  $A^{-1}BA = B$ , i.e.,

$$B^{(i)} = [A^{(i-1)}]^{-1} B^{(i-1)} A^{(i)}, \quad i = 1, 2, \dots, m-1.$$

Since  $A^{(i)}$  and  $A^{(i-1)}$  are the same similarity matrices, we thus have  $B^{(i)} = B^{(i-1)}$  for  $i = 1, 2, \dots, m-1$ , whence our theorem.

An  $m$ -adic linear group which is reducible to a 2-group automatically satisfies the condition of this theorem via its invariant first order element. An interesting property of any  $m$ -adic linear group thus conjugate to an "equi-component" group is that its  $m$ -adic collineation-group is identical with its collineation- $m$ -adic group. In fact, in the case of an equi-component group itself, the associated ordinary group consists of  $(m-1)$ -ads with equal com.

(9) Nevertheless, the set of such  $m$ -adic matrices of an  $m$ -adic linear group do constitute a subgroup, if existent, though in general not an invariant subgroup, of the group—likewise, those of these matrices having equal components. On the other hand, the subset of  $m$ -adic similarity matrices, in the sense about to be defined, while constituting an invariant subset of the  $m$ -adic linear group by their very definition, do not in general constitute a subgroup thereof. They do, however, when existent, separate into a number of semi-invariant subgroups with the subgroup of similarity- $(m-1)$ -ads as common associated group.

ponents, and hence has no other similarity- $(m-1)$ -ads than those with equal components; while under transformation by an  $(m-1)$ -ad the similarity- $(m-1)$ -ads are unchanged. An equi-component group clearly has the following two properties: (a), it is simply isomorphic with a group of ordinary matrices in the specified number of variables, (b), no two distinct elements of the group have a pair of corresponding components the same. Now these properties are invariant for transformation by an  $(m-1)$ -ad; (a), by its very formulation, (b), by our formulas for transformation by an  $(m-1)$ -ad. Hence they are satisfied by all groups conjugate to equi-component groups. The class of groups satisfying condition (a), as well as the class of groups satisfying condition (b), are therefore each at least as wide as the class of groups conjugate to equi-component groups. Actually each of the first two classes is wider than the third, for the following examples show that neither of the first two contains the other<sup>(100)</sup>. Let  $G_0$  be the axial group with elements  $((1, 1), (-1, -1)), ((-1, -1), (1, 1)), ((-1, -1), (-1, -1)), ((1, 1), (1, 1))$ ;  $S_0 = [(1, 1), (1, 1)]$ . Then in terms of the present operations the conditions of the construction theorem of §8 are satisfied, and  $G = G_0 S_0$  is a triadic linear group in two variables. Now let  $\bar{G}_0$  be the axial group with elements  $(1, -1), (-1, 1), (-1, -1), (1, 1)$ ;

$$\bar{S}_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then  $\bar{G} = \bar{G}_0 \bar{S}_0$  is a 3-group of ordinary matrices in two variables. With elements of  $G_0$  and  $\bar{G}_0$  corresponding in order,  $S_0$  corresponding to  $\bar{S}_0$ , the conditions of the simple isomorphism theorem of §8 are satisfied, so that  $G$  is simply isomorphic with  $\bar{G}$ . Hence  $G$  satisfies condition (a), but clearly fails to satisfy condition (b), since condition (b) is equivalent to the same condition stated for  $G_0$ . For our second example we consider the rather trivial case  $n=1$ . With  $G_0$  the cyclic group whose elements are  $((i), (-i)), ((-1), (-1)), ((-i), (i)), ((1), (1))$ , and  $S_0 = [(1), (1)]$ ,  $G = G_0 S_0$  is a triadic linear group in one variable satisfying condition (b). But it cannot satisfy condition (a); for it is non-abelian, while any polyadic group of ordinary matrices in one variable is readily seen to be abelian.

We conclude this section with a proof of the following generalization of the corresponding ordinary theorem. *Any abelian  $m$ -adic linear group is conjugate to a group each of whose elements is in canonical form with equal components.* We first prove this result for the case of an abelian group  $G$  having an  $m$ -adic similarity-matrix  $A$ . By the proof of the theorem preceding the above digression,  $G$  is conjugate to an equi-component group  $\bar{G}$  in which  $\bar{A}$ , the correspondent of  $A$ , has for its components equal similarity-matrices. Now the constituent  $\bar{G}'_0$  of the associated ordinary group  $\bar{G}_0$  of  $\bar{G}$  will be an ordi-

<sup>(100)</sup> Clearly these distinctions constitute but a first glance at a probably wide theory.



nary abelian linear group, and hence can be transformed by an ordinary matrix  $\alpha'$  so that each of its elements appears in canonical form. Since  $\bar{G}_0$  will consist of  $(m-1)$ -ads with equal components, the  $(m-1)$ -ad  $\alpha = (\alpha', \alpha', \dots, \alpha')$  will transform  $\bar{G}_0$  into a group in which each element appears with equal components in canonical form. As  $\alpha$  transforms  $\bar{A}$  into itself, it will therefore transform  $\bar{G} = \bar{G}_0 \bar{A}$  into the conjugate of  $G$  of our theorem.

Now let  $G$  be an arbitrary abelian  $m$ -adic linear group,  $A$  some fixed element thereof. By a previous result, we may assume the group to have been so transformed by an  $(m-1)$ -ad that  $A$  appears in canonical form with equal components  $A'$ . The  $(m-1)$ -ad  $A^{m-1}$  then has the equal components  $A'^{m-1}$ , also in canonical form. It follows from the invariance of any element  $B = [B', B'', \dots, B^{m-1}]$  of  $G$  under  $A^{m-1}$  that

$$A'^{m-1} B^{(i)} = B^{(i)} A'^{m-1}$$

for each  $i$ . If then we separate the variables of each space  $\Sigma^{(i)}$  into sets  $\Sigma_1^{(i)}, \Sigma_2^{(i)}, \dots, \Sigma_l^{(i)}$  according to their distinct multipliers in  $A'^{m-1}$ , the proof of the corresponding ordinary theorem shows that  $B^{(i)}$  transforms the variables of each  $\Sigma_j^{(i)}$  into those of  $\Sigma_j^{(i+1)}$ . Each element  $B$  of  $G$  therefore transforms  $\Sigma_j' \rightarrow \Sigma_j'' \rightarrow \dots \rightarrow \Sigma_j^{(m-1)} \rightarrow \Sigma_j'$ . That is,  $G$  appears in intransitive form with the  $l$  sets of intransitivity corresponding to  $j=1, 2, \dots, l$ . Now for each set of intransitivity the corresponding partial transformations constitute any abelian  $m$ -adic linear group. Moreover, the corresponding partial transformation of  $A$  is an  $m$ -adic similarity-matrix, since the corresponding partial transformation of  $A'^{m-1}$  has but one distinct multiplier. Hence, by our special result, each of these constituent groups can be thrown into the desired form by transformation by an  $(m-1)$ -ad on the corresponding set of intransitivity. Together, these  $l$  partial  $(m-1)$ -ads constitute an  $(m-1)$ -ad on  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  which transforms  $G$  into the conjugate group of our theorem.

Clearly, every  $m$ -adic linear group, each of whose elements is in canonical form with equal components, is abelian. On the other hand, unlike the ordinary case, an  $m$ -adic linear group each of whose elements is in canonical form need not be abelian. It is readily proved that the necessary and sufficient condition that such a group be abelian is that its associated ordinary group consist of elements with equal components.

**39.  $m$ -adic invariants.** In the theory of ordinary linear groups in  $n$  variables the concept of a function of those variables precedes that of an invariant. In our theory of  $m$ -adic linear groups  $G$  in  $n$  variables it is therefore natural to replace the concept of a function by a set of  $m-1$  functions, one for each of the spaces  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ . If we transform such a set of functions  $[f'(x_{11}, x_{12}, \dots, x_{1n}), f''(x_{21}, x_{22}, \dots, x_{2n}), \dots, f^{(m-1)}(x_{(m-1)1}, x_{(m-1)2}, \dots, x_{(m-1)n})]$  by an  $m$ -adic linear transformation  $T$  of  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$ , each function  $f^{(i)}(x_{i1}, x_{i2}, \dots, x_{in})$  will become a function of  $x_{(i+1)1}, x_{(i+1)2}, \dots, x_{(i+1)n}$ .

We therefore define  $f = [f', f'', \dots, f^{(m-1)}]$  to be an (absolute)  $m$ -adic invariant of  $T$  if  $T$  transforms  $f' \rightarrow f'', f'' \rightarrow f''', \dots, f^{(m-1)} \rightarrow f'$ ; of  $G$ , if  $f$  is an  $m$ -adic invariant of each element of  $G$ . Actually, the following analysis shows this definition to be too narrow for a real generalization of the ordinary concept. But how to widen it without destroying our basic concept of  $m-1$  spaces  $\Sigma', \Sigma'', \dots, \Sigma^{(m-1)}$  we do not at present know.

Our chief result involves the associated constituent groups  $G'_0, G''_0, \dots, G_0^{(m-1)}$  of  $G$  already introduced in §37 as the complete analogues of the corresponding concepts for  $m$ -adic substitution groups. More specifically, we saw that if  $G$  is an  $m$ -adic linear group of  $m$ -adic matrices  $T = [T', T'', \dots, T^{(m-1)}]$ ,  $G_0$ , the associated ordinary group of  $G$ , may be concretely given by a group of  $(m-1)$ -ads  $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$ . For each  $\tau$ ,  $\tau^{(i)}$  represents a transformation of the space  $\Sigma^{(i)}$  into itself; and the set of  $\tau^{(i)}$ 's constitute an ordinary group, the associated constituent group  $G_0^{(i)}$  above. It is then fundamental that, as in the case of  $m$ -adic substitution groups, the associated constituent groups of  $G$  are conjugate, each element  $T$  of  $G$  in fact transforming  $G'_0 \rightarrow G''_0, G''_0 \rightarrow G'''_0, \dots, G_0^{(m-1)} \rightarrow G'_0$ . To verify this fact we need only observe that  $T$  transforms  $G_0$  into itself; while if we follow through the operations involved in  $T^{-1}\tau T$ , we see that the  $i$ th component of the resulting  $(m-1)$ -ad is the transform of the  $(i-1)$ -st component of  $\tau$  by  $T^{(i-1)}$ .

Now let  $f = [f', f'', \dots, f^{(m-1)}]$  be an  $m$ -adic invariant of  $G$ ; that is, each element of  $G$  transforms  $f' \rightarrow f'', f'' \rightarrow f''', \dots, f^{(m-1)} \rightarrow f'$ . Each element  $\tau$  of  $G_0$  may be written as the product  $T_1 T_2 \dots T_{m-1}$  of  $m-1$  elements of  $G$ . By following through these  $m-1$  transformations we see that  $\tau$  transforms  $f'$  into itself. But  $\tau$  can operate on  $f'$  only through its first constituent  $\tau'$ . Hence each  $\tau'$  transforms  $f'$  into itself, and  $f'$  is an ordinary invariant of the associated constituent group  $G'_0$ .

Conversely, let  $f'$  be any invariant of  $G'_0$ ,  $T_0$  some element of  $G$ .  $T_0$  will transform  $f'$ , a function of the variables of  $\Sigma'$ , into a function of the variables of  $\Sigma''$ . Call this function  $f''$ , i.e.,  $f'' = (f')T_0$ . Likewise write  $f''' = (f'')T_0, \dots, f^{(m-2)} = (f^{(m-3)})T_0$ . Now  $(f^{(m-1)})T_0 = (f')T_0^{m-1}$ . Since  $f'$  is an invariant of  $G'_0$ , it will actually be transformed into itself by each element of  $G_0$ , and hence by the  $(m-1)$ -ad  $T_0^{m-1}$ . That is  $(f^{(m-1)})T_0 = f'$ , and  $f = [f', f'', \dots, f^{(m-1)}]$  is an  $m$ -adic invariant of  $T_0$ . We now show that it is also an  $m$ -adic invariant of every element  $T$  of  $G$ , that is, of  $G$ . Since  $G'_0$  is the transform of  $G'_0$  under  $T_0$ , it follows that if  $\tau''$  is any element of  $G'_0$ , then for some element  $\tau'$  of  $G'_0$ ,  $(f'')\tau'' = (f')T_0 T_0^{-1}\tau' T_0 = (f')\tau' T_0 = (f')T_0 = f''$ . Hence,  $f''$  is an invariant of  $G'_0$ , and likewise  $f'''$  of  $G'_0, \dots, f^{(m-1)}$  of  $G_0^{(m-1)}$ . Each element  $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$  of  $G_0$  will therefore transform each function  $f', f'', \dots, f^{(m-1)}$  into itself. Hence, by writing an arbitrary element  $T$  of  $G$  in the form  $\tau T_0$ , with  $\tau$  in  $G_0$ , we see that  $T$ , along with  $T_0$ , will transform  $f' \rightarrow f'', f'' \rightarrow f''', \dots, f^{(m-1)} \rightarrow f'$ .

We have thus proved the following theorem. *Given an  $m$ -adic linear group*

$G$  with first associated constituent group  $G'_0$ , then every  $m$ -adic invariant  $f = [f', f'', \dots, f^{(m-1)}]$  of  $G$  is such that  $f'$  is an ordinary invariant of  $G'_0$ ; and, conversely, every ordinary invariant  $f'$  of  $G'_0$  yields an  $m$ -adic invariant  $f = [f', f'', \dots, f^{(m-1)}]$  of  $G$ . Clearly, this correspondence between  $m$ -adic invariants of  $G$  and ordinary invariants of  $G'_0$  is 1-1. A like correspondence of course exists between the  $m$ -adic invariants of  $G$  and the ordinary invariants of  $G^{(i)}_0$  for any  $i$ .

The weakness of our concept of  $m$ -adic invariants, already apparent from this reduction to ordinary invariants, is conclusively demonstrated by a consideration of invariants as group determiners. While the groups in question will in general be infinite, no part of the above discussion involves the hypothesis of finiteness in a linear group. Suppose then that  $f = [f', f'', \dots, f^{(m-1)}]$  is an  $m$ -adic invariant of at least one  $m$ -adic linear transformation  $T_0$ , and let  $G$  be the set of all  $m$ -adic linear transformations with  $f$  as  $m$ -adic invariant. It is then readily verified that  $G$  is an  $m$ -adic linear group. By the proof of the above theorem,  $f'$  is an invariant of  $G'_0$ , and, likewise,  $f''$  of  $G''_0, \dots, f^{(m-1)}$  of  $G^{(m-1)}_0$ . If then  $\tau', \tau'', \dots, \tau^{(m-1)}$  is any selection from  $G'_0, G''_0, \dots, G^{(m-1)}_0$ , and  $\tau = (\tau', \tau'', \dots, \tau^{(m-1)})$ , then  $T = \tau T_0$  has  $f$  for  $m$ -adic invariant.  $T$  is therefore in  $G$ , and hence  $\tau$  in  $G_0$ . That is, the  $m$ -adic linear group defined by a given  $m$ -adic invariant is of that special kind in which the associated ordinary group consists of all selections, written as  $(m-1)$ -ads, that can be made from the associated constituent groups.

When the above definition is extended to relative  $m$ -adic invariant, entirely corresponding results obtain. However, by a device similar to that which gave us our  $m$ -adic alternating groups, we can enlarge somewhat the role of relative  $m$ -adic invariant as group determiner.  $f = [f', f'', \dots, f^{(m-1)}]$  will be a relative  $m$ -adic invariant of an  $m$ -adic linear transformation  $T$  if  $T$  transforms  $f$  so that  $f' \rightarrow \kappa_1 f'', f'' \rightarrow \kappa_2 f''', \dots, f^{(m-1)} \rightarrow \kappa_{m-1} f'$ , the  $\kappa$ 's being constants depending on  $T$ . Each  $T$  having  $f$  as relative  $m$ -adic invariant thus determines a  $\kappa$ -sequence. Furthermore, if  $T_1, T_2, \dots, T_m$  have  $f$  as relative  $m$ -adic invariant, so also will  $T = T_1 T_2 \dots T_m$ ; and the  $\kappa$ -sequence of  $T$  is determined by the  $\kappa$ -sequences of  $T_1, T_2, \dots, T_m$  by the same equations that connected the  $\delta$ -sequences of our alternating group theory. We are thus led to a complete  $m$ -adic  $\kappa$ -group; and corresponding to any subgroup thereof, the set of all  $T$ 's with  $\kappa$ -sequences in that subgroup will be an  $m$ -adic linear group. Furthermore, whenever the associated ordinary group of the  $\kappa$ -subgroup does not consist of all selections from its constituent associated subgroups, the corresponding  $m$ -adic linear group will also not be of this special type. However, with the  $f^{(i)}$ 's homogeneous polynomials in the corresponding variables, any  $T$  having  $f$  for relative  $m$ -adic invariant can be changed to a  $T$  having  $f$  for absolute  $m$ -adic invariant by multiplying it into a suitable similarity- $(m-1)$ -ad; and conversely, without qualification. Hence the  $T$ 's corresponding to any one  $\kappa$ -sequence represent the same  $m$ -adic collineations as the

$T$ 's having  $f$  for absolute invariant. All the  $m$ -adic linear groups corresponding to the various  $\kappa$ -subgroups therefore have the same corresponding  $m$ -adic collineation-group as the  $G$  defined by  $f$  as absolute invariant, and our seemingly greater freedom is largely illusory.

An obvious, but probably superficial, remedy for the relative triviality of our concept of  $m$ -adic invariant would be to allow each of the functions  $f', f'', \dots, f^{(m-1)}$  to be functions not of the variables of the corresponding  $\Sigma$  alone, but of all of the  $\Sigma$ 's. It may be mere prejudice that makes us object to thus uniting the  $m-1$  spaces of  $n$  dimensions each into one space of  $(m-1)n$  dimensions; for, certainly, arbitrarily to give  $m-1$  points, one for each space, is equivalent to giving one point in the combined space. One qualification does suggest itself. Corresponding to the condition of homogeneity for the polynomial invariants of ordinary theory, §36 suggests that the  $f^{(i)}$ 's be polynomials homogeneous in the variables of each  $\Sigma$  separately. However, a finally acceptable form for a general concept of  $m$ -adic invariant will probably involve changes in our original idea both more specific and more drastic than here suggested.

40. **Generalization of  $m$ -adic substitution and transformation groups.** The concept of  $m$ -adic linear group is readily extended to that of an  $(m, \mu)$  linear group, analogous to our earlier  $(m, \mu)$  substitution group. However, both concepts admit of a far wider extension. We shall give this extension only for  $m$ -adic substitution groups, the generalization of  $m$ -adic linear group being entirely similar<sup>(101)</sup>. It is of interest to note that this generalization continues to be a generalization even when  $m=2$ . But the resulting ordinary groups are then essentially realizations of Specht groups, referred to in the introduction, or subgroups thereof<sup>(102)</sup>.

The concepts of an  $m$ -adic substitution on the letters of classes  $\Gamma_1, \Gamma_2, \dots, \Gamma_{m-1}$  is associated with the cyclic substitution  $(\Gamma_1 \Gamma_2 \dots \Gamma_{m-1})$  on the classes themselves; for, under the  $m$ -adic substitution,  $\Gamma_1 \rightarrow \Gamma_2, \Gamma_2 \rightarrow \Gamma_3, \dots, \Gamma_{m-1} \rightarrow \Gamma_1$ . More generally then let  $\Gamma_1, \Gamma_2, \dots, \Gamma_m$  be any finite set of classes,  $\sigma$  any substitution on those classes themselves as elements.  $s$  will then be said to be a *polyadic substitution corresponding to  $\sigma$*  if, whenever  $\sigma$  replaces class  $\Gamma_i$  by class  $\Gamma_j$ ,  $s$  carries the members of  $\Gamma_i$  in 1-1 fashion into the members of  $\Gamma_j$ . Clearly, if polyadic substitutions  $s_1, s_2, \dots, s_m$  on the members of  $\Gamma_1, \Gamma_2, \dots, \Gamma_m$  correspond to  $\sigma_1, \sigma_2, \dots, \sigma_m$  respectively,  $s_1 s_2 \dots s_m$ , the result of performing these  $m$  polyadic substitutions in succession, is itself a

<sup>(101)</sup> A corresponding generalization of our narrow concept of  $m$ -adic invariant immediately suggests itself.

<sup>(102)</sup> On the other hand, groups of the permutations of sets of variables considered by L. Weisner (*Generalization of Lagrange's theorem*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 629-630) are but a very special case of the present generalization with  $m=2$ . We may note that the associated and containing ordinary groups of  $m$ -adic substitution groups, and, indeed, of the present generalization thereof, also come under this generalization with  $m=2$ , and thus tie up with Specht groups, or subgroups thereof.

polyadic substitution corresponding to  $\sigma_1\sigma_2 \cdots \sigma_m$ , the product of the  $m$  corresponding ordinary substitutions. It follows from our last result on homomorphisms given in §4 that if  $G$  is an  $m$ -group of polyadic substitutions  $s$  on the members of  $\Gamma_1, \Gamma_2, \dots, \Gamma_r$  under the above  $m$ -adic operation, the corresponding ordinary substitutions  $\sigma$  form an  $m$ -group  $B$  of ordinary substitutions. Moreover,  $G$  is homomorphic to  $B$ . We shall call  $B$  the *basic  $m$ -group* corresponding to the *polyadic substitution group*  $G$ . In the case of our  $m$ -adic substitution groups, and more generally our  $(m, \mu)$  groups, the basic  $m$ -group is of first order, its sole substitution consisting of a single cycle the number of whose letters is  $m-1$  in the first case, a divisor  $\mu-1$  of  $m-1$  in the second.

As a consequence of the homomorphism between an arbitrary polyadic substitution group  $G$  and its basic  $m$ -group  $B$ , we see that there are the same number of polyadic substitutions in  $G$  for each substitution in  $B$ . Hence, also, the order of  $G$  is always a multiple of the order of  $B$ . Again, the ordinary substitutions corresponding to the polyadic substitutions forming any subgroup of  $G$  will form a subgroup of  $B$ , if not  $B$  itself; while to each subgroup of  $B$  there is at least one corresponding subgroup of  $G$ , i.e., the one consisting of all the elements of  $G$  corresponding to the elements of the subgroup of  $B$ , and hence containing all such subgroups.

For simplicity, we now restrict ourselves to mutually exclusive classes  $\Gamma_1, \Gamma_2, \dots, \Gamma_r$  of the same finite number of letters  $n$  each<sup>(10)</sup>. Given any substitution  $\sigma$  on those classes as elements, there will then be a total of  $(n!)^r$  polyadic substitutions corresponding to  $\sigma$ . If then  $B$  is a given  $m$ -group of substitutions on those classes as elements, and  $b$  is the order of  $B$ , the  $(n!)^rb$  polyadic substitutions corresponding to the elements of  $B$  are readily seen to constitute a polyadic substitution group with  $B$  as basic group. It may be called the  *$m$ -adic symmetric group of degree  $n$  with basic  $m$ -group  $B$* . We can now state that any polyadic group with basic  $m$ -group  $B$  is a subgroup of the corresponding  $m$ -adic symmetric group. On the other hand, a subgroup of that  $m$ -adic symmetric group may have but a subgroup of  $B$  for basic group.

Of the theory of  $m$ -adic substitution groups we shall redevelop here only the general aspects of the theory leading to  $m$ -adic alternating groups. Again form the Vandermonde determinants  $\Delta_1, \Delta_2, \dots, \Delta_r$  for the letters of  $\Gamma_1, \Gamma_2, \dots, \Gamma_r$  respectively. If now a substitution  $\sigma$  on the  $\Gamma$ 's as elements be written in the primitive form

$$\Gamma_1 \Gamma_2 \cdots \Gamma_r,$$

$$\Gamma_{i_1} \Gamma_{i_2} \cdots \Gamma_{i_r},$$

a polyadic substitution corresponding to  $\sigma$  will transform the  $\Delta$ 's as follows:

$$\Delta_1 \rightarrow \delta' \Delta_{i_1}, \Delta_2 \rightarrow \delta'' \Delta_{i_2}, \dots, \Delta^{(r)} \rightarrow \delta^{(r)} \Delta_{i_r}, \quad \delta', \delta'', \dots, \delta^{(r)} = \pm 1.$$

<sup>(10)</sup> When  $B$  is transitive, the number of letters in the several  $\Gamma$ 's must of necessity be the same.



To describe this transformation completely, we must therefore not only specify the  $\delta$ -sequence  $\delta = [\delta', \delta'', \dots, \delta^{(n)}]$ , but the substitution  $\sigma$ . We therefore form the couple  $\{\sigma, \delta\}$ . Given then a polyadic substitution group  $G$ , each element thereof uniquely determines a  $\{\sigma, \delta\}$  couple. Moreover, if  $s_1, s_2, \dots, s_m$  are any  $m$  elements of  $G$ ,  $\{\sigma_1, \delta_1\}, \{\sigma_2, \delta_2\}, \dots, \{\sigma_m, \delta_m\}$  the corresponding couples, then  $s = s_1 s_2 \dots s_m$  has a couple  $\{\sigma, \delta\}$  completely determined by the couples of  $s_1, s_2, \dots, s_m$ . For clearly  $\sigma = \sigma_1 \sigma_2 \dots \sigma_m$ . On the other hand, let  $\delta = [\delta', \delta'', \dots, \delta^{(n)}]$ ,  $\delta_i = [\delta'_i, \delta''_i, \dots, \delta^{(n)}_i]$ . For any substitution  $\sigma$  on the  $\Gamma$ 's as elements, if  $\sigma$  carries  $\Gamma_i$  into  $\Gamma_{i_j}$ , write  $i_j = i\sigma$ . Then we will have

$$\delta^{(i)} = \delta_1^{(i)} \delta_2^{(i\sigma_1)} \dots \delta_{m-1}^{(i\sigma_1 \dots \sigma_{m-2})} \delta_m^{(i\sigma_1 \dots \sigma_{m-2} \sigma_{m-1})}.$$

It again follows from our last result on homomorphisms that the class of  $\{\sigma, \delta\}$  couples corresponding to the elements of  $G$  constitutes an  $m$ -group under the resulting  $m$ -adic operation on  $\{\sigma, \delta\}$  couples, and hence that  $G$  is homomorphic to this  $m$ -group. We shall call the latter the  $\{\sigma, \delta\}$  subgroup corresponding to  $G$ . The homomorphism in question then again assures us that there are exactly the same number of elements of  $G$  for each  $\{\sigma, \delta\}$  couple in its  $\{\sigma, \delta\}$  subgroup, and again yields the many-one relation between the subgroups of  $G$  and those of its  $\{\sigma, \delta\}$  subgroup.

Clearly the relationship between  $G$  and its  $\{\sigma, \delta\}$  subgroup is intimately bound up with the relationship between  $G$  and its basic  $m$ -group  $B$ . In fact, the very form of a  $\{\sigma, \delta\}$  couple yields a many-one correspondence between the elements of the  $\{\sigma, \delta\}$  subgroup corresponding to  $G$ , and of  $B$ ; while our formulation of the  $m$ -adic operation on  $\{\sigma, \delta\}$  couples shows this correspondence to be a homomorphism—hence again the sameness of the number of  $\{\sigma, \delta\}$  couples corresponding to different  $\sigma$ 's, and the many-one correspondence between the subgroups of the  $\{\sigma, \delta\}$  subgroup, and of the basic  $m$ -group  $B$ , corresponding to  $G$ . Much can now be said of the interrelations between  $G$ , its  $\{\sigma, \delta\}$  subgroup, and its basic  $m$ -group  $B$ . But they are all implicit in the fact that the above homomorphism between  $G$  and  $B$  is the one determined by the homomorphism between  $G$  and its  $\{\sigma, \delta\}$  subgroup, and the homomorphism between that  $\{\sigma, \delta\}$  subgroup and  $B$ .

When  $G$  is the polyadic symmetric group of degree  $n$  corresponding to a given basic  $m$ -group  $B$ , then, as in the case of  $m$ -adic substitutions,  $G$  will have at least one polyadic substitution for each of the  $2^n$  possible  $\delta$ -sequences, and each substitution  $\sigma$  in  $B$ , provided  $n > 1$ . The " $\{\sigma, \delta\}$  subgroup" may now be called the *complete*  $\{\sigma, \delta\}$  group corresponding to  $B$ . With  $B$  of order  $b$ , the corresponding complete  $\{\sigma, \delta\}$  group is then of order  $2^n b$ . We thus have a division of the corresponding  $(n!)^b$  polyadic substitutions into  $2^n b$  mutually exclusive classes of consequently  $(n!/2)^b$  members each.

Now in the many-one relations between the subgroups of the polyadic symmetric group of degree  $n$ , the complete  $\{\sigma, \delta\}$  group, and the basic

$m$ -group  $B$  consider only those (proper) subgroups of the complete  $\{\sigma, \delta\}$  group which correspond to  $B$  itself. For each of these  $\{\sigma, \delta\}$  subgroups there is a unique largest subgroup of the polyadic symmetric group. These may then be called the *polyadic alternating groups* of degree  $n$  with basic  $m$ -group  $B$ . The corresponding  $\{\sigma, \delta\}$  subgroups are of orders  $2^{\nu_1 b}$ ,  $0 \leq \nu_1 < \nu$ , and the polyadic alternating groups correspondingly of orders  $(n!/2)^{\nu_1 b}$ , each consisting of all the elements in each of  $2^{\nu_1 b}$  of the above mutually exclusive classes. Note that if  $B$  is considered as a substitution group on the symbols  $\Gamma_1, \Gamma_2, \dots, \Gamma$ , rather than on the classes they symbolize, then one and the same  $B$  will serve for arbitrary  $n$ . Hence also the complete  $\{\sigma, \delta\}$  group will be independent of  $n$ ; and for each  $n > 1$  there will be as many polyadic alternating groups of degree  $n$  and basic  $m$ -group  $B$  as the complete  $\{\sigma, \delta\}$  group has subgroups also corresponding to  $B$ .

By considering an arbitrary polyadic group  $G$  of degree  $n$ , and with basic  $m$ -group  $B$ , a subgroup of the corresponding polyadic symmetric group, we see that the  $\{\sigma, \delta\}$  subgroup for  $G$  is actually a subgroup, proper or improper, of the complete  $\{\sigma, \delta\}$  group corresponding to  $B$ . But that subgroup also must correspond to  $B$ . That is, we have a many-one relation between all polyadic groups of degree  $n$  with basic  $m$ -group  $B$ , and those subgroups of the complete  $\{\sigma, \delta\}$  group which themselves correspond to  $B$ .

COLLEGE OF THE CITY OF NEW YORK,  
NEW YORK, N.Y.

# ON A MINIMUM PROBLEM IN THE THEORY OF ANALYTIC FUNCTIONS OF SEVERAL VARIABLES

BY

W. T. MARTIN

1. **Introduction.** In 1932 Wirtinger<sup>(1)</sup> posed and solved the following problem. Given a region  $G$  in the complex  $z$ -plane and a (complex-valued) function  $\phi(z, \bar{z}) \equiv \phi(x+iy, x-iy)$  continuous and with continuous first partial derivatives with respect to  $x$  and  $y$  in  $G$ , to find an analytic function  $f(z)$  which gives the best approximation to  $\phi$  in the mean-square sense, that is, such that

$$(1.1) \quad \int_G |\phi - f|^2 d\omega_z = \min,$$

where  $d\omega_z$  is the element of area  $dx dy$ . (In the case in which  $G$  extends to infinity he also assumed that  $\phi \in L^2$  over  $G$ .) By use of the Green's function  $G(z, \bar{z}; \zeta, \bar{\zeta})$  for the region  $G$ , he proved the existence (and uniqueness) of such an  $f$  and gave an explicit formula for  $f$ , namely,

$$(1.2) \quad f(z) = \frac{1}{2\pi} \int_G \phi(\zeta, \bar{\zeta}) \frac{\partial^2 G}{\partial z \partial \bar{\zeta}} d\omega_{\zeta},$$

when  $\partial/\partial z \equiv \frac{1}{2}(\partial/\partial x - i\partial/\partial y)$ ,  $\partial/\partial \bar{z} \equiv \frac{1}{2}(\partial/\partial x + i\partial/\partial y)$ , etc. In the case of the unit circle  $C$  the formula yields the result

$$(1.3) \quad f(z) = -\frac{1}{\pi} \int_C \frac{\phi(\zeta, \bar{\zeta})}{(1 - z\bar{\zeta})^2} d\omega_{\zeta},$$

a result which he also obtained directly by use of the Fourier series for  $\phi$  and  $f$ .

Recently Wirtinger<sup>(2)</sup> posed the analogous question in the theory of functions of several complex variables. In the case in which the region under consideration is a hypersphere  $H = E[|z_1|^2 + \dots + |z_n|^2 < 1]$  and  $\phi(z_1, \dots, z_n; \bar{z}_1, \dots, \bar{z}_n)$  is merely integrable over  $H$ , he obtained a (unique) solution by use of multiple Fourier series, namely

$$(1.4) \quad f(z_1, \dots, z_n) = (-1)^n \frac{n!}{\pi^n} \int_H \frac{\phi(\zeta_1, \dots, \zeta_n; \bar{\zeta}_1, \dots, \bar{\zeta}_n)}{[1 - (z_1 \bar{\zeta}_1 + \dots + z_n \bar{\zeta}_n)]^{n+1}} d\omega_{\zeta},$$

Presented to the Society, October 28, 1939; received by the editors January 22, 1940. This paper was received by the editors of the Bulletin of the American Mathematical Society September 26, 1939, accepted by them, and later transferred to these Transactions.

(<sup>1</sup>) W. Wirtinger, Monatshefte für Mathematik und Physik, vol. 39 (1932), pp. 377-384.

(<sup>2</sup>) W. Wirtinger, Monatshefte für Mathematik und Physik, vol. 47 (1939), pp. 426-431.

where  $d\omega_r$  is the  $2n$ -dimensional volume element. He conjectured that the question probably has a solution for general regions and for very general functions  $\phi$  but that the solution appeared to involve difficult investigations on the extensions of Green's functions. Now in various questions in the theory of functions of several complex variables Bergmann has been able to replace the theory of the Green's functions by the theory of complex orthogonal functions and the kernel of a region<sup>(3)</sup>. In this note we show that by the use of this theory of the kernel of a region we can solve the problem posed by Wirtinger for a very general class of regions (which includes all bounded regions) and for  $\phi$  belonging to  $L^2$ ; indeed, we give the solution explicitly in terms of an integral involving  $\phi$  and the kernel of the region (see equation (3.10)).

It is known that results of this general nature have important applications; for example in connection with the theory of entire functions of two variables Bergmann has solved the same problem with the function  $f$  bi-harmonic (the real part of an analytic function of two variables) rather than analytic<sup>(4)</sup>.

For the sake of completeness we shall give in §2 a brief résumé of the results from the theory of orthogonal functions and the kernel of a region. Also in the concluding section we consider certain extensions of the problem. We shall speak only of two variables; the case of  $n$  variables involves no essential changes.

**2. The kernel of a region.** To every region of a wide class of four-dimensional regions there corresponds a kernel function which is defined as follows<sup>(5)</sup>. Let  $\mathcal{B}$  be a region of this class and let  $\{\Omega^{(\nu)}(z_1, z_2)\}$  be a complete orthonormal system of analytic functions belonging to  $L^2$  over  $\mathcal{B}$ , so that

$$(2.1) \quad \int_{\mathcal{B}} \Omega^{(\nu)}(z_1, z_2) \overline{\Omega^{(\mu)}(z_1, z_2)} d\omega_z = \delta_{\mu\nu}, \quad \mu, \nu = 1, 2, \dots,$$

where  $\int_{\mathcal{B}} = \lim_{m \rightarrow \infty} \int_{\mathcal{B}_m}$  and  $\{\mathcal{B}_m\}$  is a system of regions in  $\mathcal{B}$  converging to  $\mathcal{B}$ . The series

$$(2.2) \quad \sum_{\nu=1}^{\infty} \Omega^{(\nu)}(z_1, z_2) \overline{\Omega^{(\nu)}(\xi_1, \xi_2)}$$

<sup>(3)</sup> For the development of the theory of the kernel of a region see S. Bergmann, *Mathematische Zeitschrift*, vol. 29 (1929), pp. 640-677, *Journal für die reine und angewandte Mathematik*, vol. 169 (1933), pp. 1-42, especially pp. 1-5; vol. 172 (1934), pp. 89-128. We shall refer to these papers as  $B_1$  and  $B_2$  respectively.

<sup>(4)</sup> S. Bergmann, *Mathematische Annalen*, vol. 109 (1934), pp. 324-348, especially p. 333; *Compositio Mathematica*, vol. 3 (1934), pp. 137-173. We shall refer to these papers as  $B_3$  and  $B_4$  respectively.

<sup>(5)</sup> The results which we state in this section are all given by Bergmann in the papers listed in footnotes 3 and 4. We shall restrict ourselves to simply-connected bounded regions but the results are true for any region for which there exists a set of linearly independent functions belonging to  $L^2$ .

converges absolutely and uniformly for  $(z)$  and  $(\bar{z})$  in any regions interior to  $\mathcal{B}$  and accordingly defines a function of  $z_1, z_2, \bar{z}_1, \bar{z}_2$  analytic for  $(z)$  and  $(\bar{z})$  in  $\mathcal{B}$  (see  $B_2$ ). The sum function is called the kernel of the region  $\mathcal{B}$  and is denoted by  $K_{\mathcal{B}}(z_1, z_2, \bar{z}_1, \bar{z}_2)$ . It is known that the function depends only upon the region  $\mathcal{B}$  and not upon the particular set of orthonormal functions used in defining it (see  $B_2$ ). Concerning series in terms of the  $\Omega^{(v)}$  it has been shown that the series

$$(2.3) \quad \left| \sum_{v=1}^{\infty} a_v \Omega^{(v)}(z_1, z_2) \right|^2$$

can be integrated term-by-term over  $\mathcal{B}$  whenever  $\sum |a_v|^2 < \infty$  (see  $B_3$ , p. 331).

3. **Solution of the problem.** Let  $\phi(z_1, z_2; \bar{z}_1, \bar{z}_2)$  be a complex-valued function of the four real variables  $x_1, x_2, y_1, y_2$ , defined and of integrable square over a bounded region  $\mathcal{B}$

$$(3.1) \quad \int_{\mathcal{B}} |\phi|^2 d\omega_s < \infty.$$

We seek a function  $f(z_1, z_2)$  analytic and of integrable square over  $\mathcal{B}$  and such that

$$(3.2) \quad \int_{\mathcal{B}} |\phi - f|^2 d\omega_s = \min.$$

For the solution let  $\{\Omega^{(v)}(z_1, z_2)\}$  be a complete orthonormal set of analytic functions belonging to  $L^2$  over  $\mathcal{B}$  and let us seek to determine coefficients  $\{a_v\}$  subject to the condition

$$(3.3) \quad \sum_1^{\infty} |a_v|^2 < \infty$$

in such a manner that the function

$$(3.4) \quad f(z_1, z_2) = \sum_1^{\infty} a_v \Omega^{(v)}(z_1, z_2)$$

furnishes a minimum to (3.2). If we substitute (3.4) into (3.2) we find

$$(3.5) \quad \begin{aligned} \int_{\mathcal{B}} |\phi - f|^2 d\omega &= \int_{\mathcal{B}} |\phi|^2 d\omega - \int_{\mathcal{B}} \left( \sum_1^{\infty} a_v \Omega^{(v)} \right) \bar{\phi} d\omega \\ &\quad - \int_{\mathcal{B}} \left( \sum_1^{\infty} \bar{a}_v \overline{\Omega^{(v)}} \right) \phi d\omega + \int_{\mathcal{B}} \left| \sum_1^{\infty} a_v \Omega^{(v)} \right|^2 d\omega. \end{aligned}$$

Using (3.3) and the results stated in §2 we see that we may integrate term-wise, thus



$$(3.6) \quad \int_{\mathfrak{B}} |\phi - f|^2 d\omega = \int_{\mathfrak{B}} |\phi|^2 d\omega - \sum_1^{\infty} (a_r \bar{b}_r + \bar{a}_r b_r - a_r \bar{a}_r),$$

where we have written

$$(3.7) \quad b_r = \int_{\mathfrak{B}} \phi \bar{\Omega}^{(r)} d\omega.$$

By Bessel's inequality

$$(3.8) \quad \sum_{r=1}^{\infty} |b_r|^2 < \infty.$$

Treating  $a_r, \bar{a}_r$  as independent complex variables and differentiating with respect to  $\bar{a}_r$  (or  $a_r$ ) we see that Euler's conditions for (3.6) to be a minimum are

$$(3.9) \quad a_r = b_r, \quad r = 1, 2, \dots$$

Clearly this choice of the  $a$ 's furnishes an actual minimum (we shall also give a direct proof of this fact in equation (3.12) below). Thus the minimizing function  $f$  has the form

$$\begin{aligned} f(z_1, z_2) &= \sum_1^{\infty} \Omega^{(r)}(z_1, z_2) \int_{\mathfrak{B}} \phi(\xi_1, \xi_2; \bar{\xi}_1, \bar{\xi}_2) \overline{\Omega^{(r)}(\xi_1, \xi_2)} d\omega_{\xi} \\ &= \int_{\mathfrak{B}} \phi(\xi_1, \xi_2; \bar{\xi}_1, \bar{\xi}_2) \sum_1^{\infty} \Omega^{(r)}(z_1, z_2) \overline{\Omega^{(r)}(\xi_1, \xi_2)} d\omega_{\xi} \\ &= \int_{\mathfrak{B}} \phi(\xi_1, \xi_2; \bar{\xi}_1, \bar{\xi}_2) K_{\mathfrak{B}}(z_1, z_2, \bar{\xi}_1, \bar{\xi}_2) d\omega_{\xi}, \end{aligned}$$

where we have again used the fact that we may interchange the order of integration and summation.

Thus we have answered Wirtinger's question.

**THEOREM.** Let  $\mathfrak{B}$  be any (four-dimensional) region for which there exists an infinite system of linearly independent analytic functions of  $L^2$ . (In particular, let  $\mathfrak{B}$  be any simply connected bounded region.) Let  $\phi(z_1, z_2, \bar{z}_1, \bar{z}_2)$  be of integrable square over  $\mathfrak{B}$ . Then the function  $f$  defined by

$$(3.10) \quad f(z_1, z_2) = \int_{\mathfrak{B}} \phi(\xi_1, \xi_2; \bar{\xi}_1, \bar{\xi}_2) K_{\mathfrak{B}}(z_1, z_2, \bar{\xi}_1, \bar{\xi}_2) d\omega_{\xi},$$

where  $K_{\mathfrak{B}}$  is the kernel of the region  $\mathfrak{B}$  defined as in (2.2), is analytic and of integrable square over  $\mathfrak{B}$  and furnishes the unique minimum to the integral (3.2) over the class of analytic functions of integrable square.

Very many different properties of the kernel function are known which

yield various properties of the minimizing function  $f$ ; for example if  $g(z_1, z_2)$  is any analytic function of  $L^2$  over  $\mathcal{B}$  then<sup>(6)</sup>

$$(3.11) \quad \int_{\mathcal{B}} (\phi - f) \bar{g} d\omega_s = 0.$$

This result is the analogue for the region  $\mathcal{B}$  of a result obtained by Wirtinger for the hypersphere (loc. cit., footnote 2, equation (8)). It also obviously furnishes a direct proof of the fact that the function  $f$  defined in (3.10) yields a minimum for (3.2), since in view of (3.11), if  $g \neq 0$ ,

$$(3.12) \quad \begin{aligned} \int |f + g - \phi|^2 &= \int |f - \phi|^2 + \int (f - \phi) \bar{g} + \int (\bar{f} - \bar{\phi}) g + \int |g|^2 \\ &= \int |f - \phi|^2 + \int |g|^2 > \int |f - \phi|^2. \end{aligned}$$

4. **Special regions.** For many special regions the kernel function has been given explicitly, for instance in the case of a Reinhardt region in four-dimensional space

$$(4.1) \quad R = E[|z_2|^2 < G(|z_1|^2), \quad 0 \leq |z_1| < 1],$$

where  $G$  is once differentiable in  $(0, 1)$ , the kernel has the form (see B<sub>1</sub>)

$$(4.2) \quad K_R(z_1, z_2, \bar{z}_1, \bar{z}_2) = \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} \frac{z_1^{m-m} \bar{z}_1^{p-p} z_2^m \bar{z}_2^p}{[\pi^2/(p+1)] \int_0^1 \rho^m [G(\rho)]^{p+1} d\rho}.$$

If we have a region  $R^*$  which can be mapped into  $R$  by means of a transformation  $z_\kappa = z_\kappa(w_1, w_2)$ ,  $\kappa = 1, 2$ , where the  $z_\kappa$  are analytic in  $R^*$ , then the kernel function for  $R^*$  is equal to the kernel for  $R$  multiplied by the two jacobians of the transformation (see B<sub>1</sub>, p. 5):

$$(4.3) \quad K_{R^*}(w_1, w_2, \bar{w}_1, \bar{w}_2) = \left[ K_R(z_1(w_1, w_2), z_2(w_1, w_2), \bar{z}_1(\bar{w}_1, \bar{w}_2), \bar{z}_2(\bar{w}_1, \bar{w}_2)) \right. \\ \left. \frac{D(z_1, z_2)}{D(w_1, w_2)} \frac{\overline{D(z_1, z_2)}}{\overline{D(\bar{w}_1, \bar{w}_2)}} \right].$$

In different cases the series in (4.2) can be summed, for example in the

(<sup>6</sup>) We may see this fact directly in view of the form of  $f$  and the orthogonality of the  $\Omega$ 's, or we may note that the corresponding result for the case of biharmonic functions has been proved by Bergmann (see B<sub>2</sub>, p. 333). In order to see it directly let us write  $c_r = f g \Omega^{(r)}$ . Then  $g(z_1, z_2) = \sum_1^{\infty} c_r \Omega^{(r)}(z_1, z_2)$ . Also by (3.4), (3.7) and (3.9)  $f(z_1, z_2) = \sum_1^{\infty} b_r \Omega^{(r)}(z_1, z_2)$  where  $b_r = f \phi \Omega^{(r)}$ . Thus

$$\begin{aligned} f(\phi - f) \bar{g} &= f \phi \sum_1^{\infty} \bar{c}_r \overline{\Omega^{(r)}} - f [\sum_1^{\infty} b_r \Omega^{(r)}] [\sum_1^{\infty} \bar{c}_r \overline{\Omega^{(r)}}] \\ &= \sum_1^{\infty} \bar{c}_r f \phi \overline{\Omega^{(r)}} - \sum_1^{\infty} \bar{c}_r b_r = 0. \end{aligned}$$

case of a region of the form

$$(4.4) \quad a |z_1|^{2/p} + |z_2|^2 < 1, \quad p \text{ integral, } p > 0, 0 < a \leq 1,$$

the kernel is (see B<sub>1</sub>)

$$(4.5) \quad K(z_1, z_2; \bar{z}_1, \bar{z}_2) = a^p (1 - z_2 \bar{z}_2)^{p-2} \frac{(p+1)(1 - z_2 \bar{z}_2)^p + (p-1)a^p z_1 \bar{z}_1}{\pi^2 [(1 - z_2 \bar{z}_2)^p - a^p z_1 \bar{z}_1]^3}$$

which yields for the hypersphere  $H = E[|z_1|^2 + |z_2|^2 < 1]$  the result

$$(4.6) \quad K_H(z_1, z_2; \bar{z}_1, \bar{z}_2) = \frac{2}{\pi^2 [1 - z_1 \bar{z}_1 - z_2 \bar{z}_2]^3}.$$

If we put this into (3.10), then we see that for the hypersphere  $H$  our result is identical with the formula (1.3) obtained by Wirtinger (for  $n=2$ ).

It is perhaps worth while merely to mention that in the case of a bicylinder  $|z_k| < r_k$ ,  $k=1, 2$ , the kernel has the form (see B<sub>1</sub>)

$$K(z_1, z_2, \bar{z}_1, \bar{z}_2) = \frac{r_1^2 r_2^2}{\pi^2 (r_1^2 - z_1 \bar{z}_1)^2 (r_2^2 - z_2 \bar{z}_2)^2}.$$

It is also interesting that in the case of simply connected regions in the complex  $z$ -plane the kernel is simply the expression  $\partial^2 G(z, \bar{z}) / \partial z \partial \bar{z}$  where  $G$  is the Green's function for the region<sup>(7)</sup>. This of course is in agreement with the result (1.2) of Wirtingers' mentioned in the introduction.

**5. Extensions.** A very important variation of the problem in the theory of functions of one complex variable is the case in which the integration is over the boundary curve. In the case of two complex variables, when the region under consideration has a *distinguished boundary surface*, the analogous problem may be solved and since there is a general theory of orthogonal functions and kernel functions related to the distinguished boundary surface<sup>(8)</sup>, the same formula for  $f$  as in (3.10) is obtained, with of course the kernel  $K$  defined analogously.

Moreover we may ask not only that  $f$  be analytic and of  $L^2$  over  $\mathcal{B}$  and minimize the integral (3.2) but also that  $f$  be subjected to certain additional conditions, for example that

$$(5.1) \quad f(t_1^{(s)}, t_2^{(s)}) = X_s, \quad s = 1, \dots, p,$$

<sup>(7)</sup> The kernel for doubly connected regions in the complex  $z$ -plane has been calculated by K. Zarankiewicz, *Zeitschrift für angewandte Mathematik und Mechanik*, vol. 14 (1934), pp. 97-104 and by P. Kufareff, *Bulletin de l'Institut Mathématique et Mécanique*, Tomsk, vol. 1 (1937), pp. 228-235.

<sup>(8)</sup> See Bergmann, *Bulletin de l'Institut Mathématique et Mécanique*, Tomsk, vol. 3 (1935-1937), pp. 242-257.

where  $\{t_1^{(v)}, t_2^{(v)}\} \in \mathcal{B}$ . We shall merely indicate the proof in the case  $p=1$ . Our problem is then to find an  $f$  analytic and of  $L^2$  over  $\mathcal{B}$ , which minimizes the integral (3.2) and which takes on a given value  $X$  at a fixed point  $(t_1, t_2)$  in  $\mathcal{B}$ ,

$$(5.2) \quad f(t_1, t_2) = X.$$

The analogue of (3.6) is

$$\begin{aligned} \int |\phi - f|^2 - \lambda[f(t) - X] - \mu[\overline{f(t)} - \overline{X}] \\ = \int |\phi|^2 - \sum_1^\infty (a_v \bar{b}_v + a_v b_v - a_v \bar{a}_v) - \lambda \left[ \sum_1^\infty a_v \Omega^{(v)}(t) - X \right] \\ - \mu [\sum \bar{a}_v \bar{\Omega}^{(v)}(t) - \bar{X}] \end{aligned}$$

where  $\lambda, \mu$  are the Lagrangian multipliers. Euler's conditions are

$$(5.3) \quad a_v = b_v + \mu \overline{\Omega^{(v)}(t)}, \quad \bar{a}_v = \bar{b}_v + \lambda \Omega^{(v)}(t), \quad v = 1, 2, \dots$$

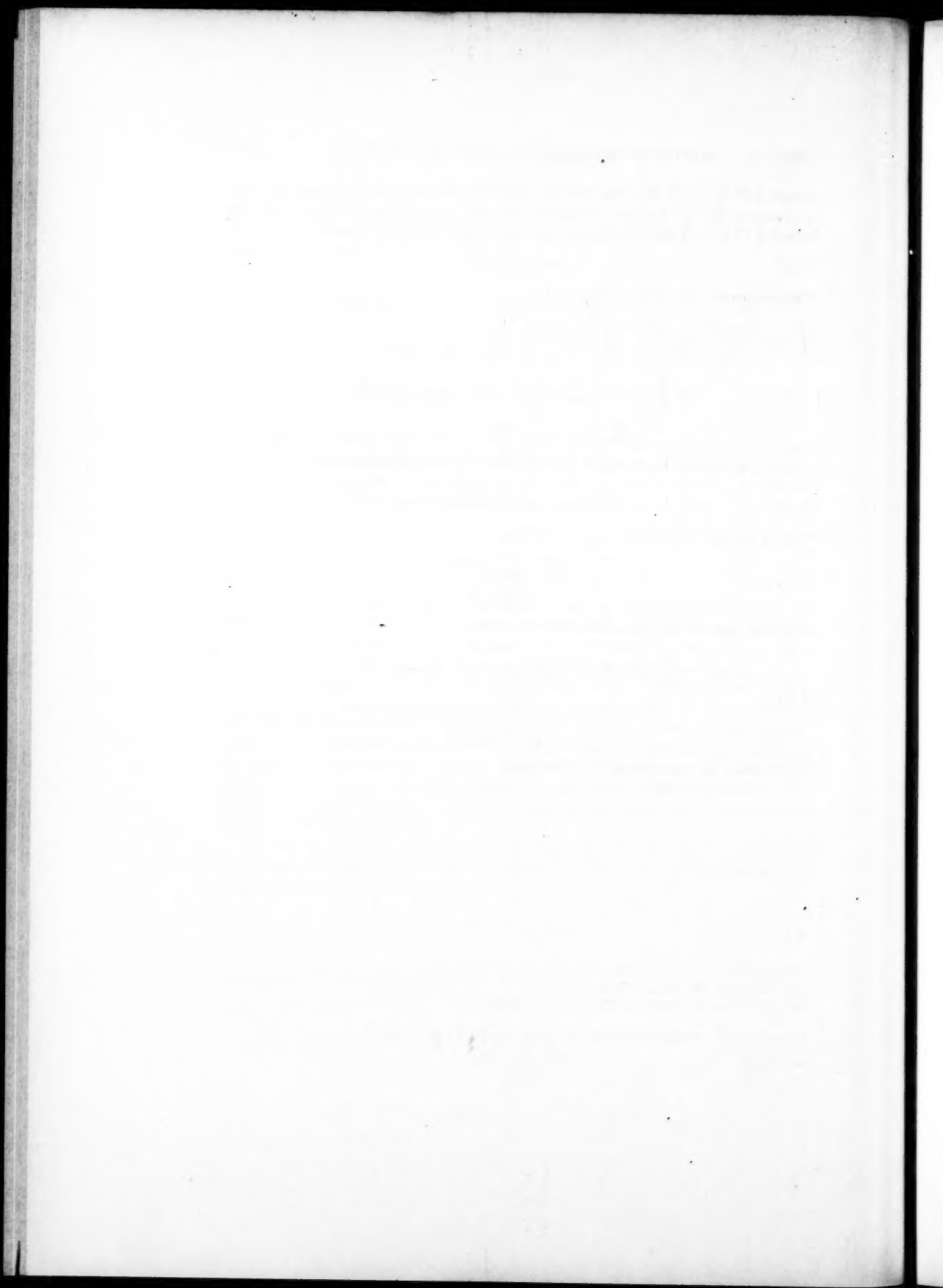
Thus  $\mu = \bar{\lambda}$  and the condition (5.2) yields

$$(5.4) \quad \mu = \frac{X - \sum_1^\infty b_v \Omega^{(v)}(t)}{K_{\mathcal{B}}(t, \bar{t})}.$$

Thus the minimizing function  $f$  has the form

$$\begin{aligned} f(z_1, z_2) = \int_{\mathcal{B}} \phi(\xi_1, \xi_2; \bar{\xi}_1, \bar{\xi}_2) K_{\mathcal{B}}(z_1, z_2; \bar{\xi}_1, \bar{\xi}_2) d\omega_{\xi} \\ (5.5) \quad + \frac{X - \int_{\mathcal{B}} \phi(\xi_1, \xi_2; \bar{\xi}_1, \bar{\xi}_2) K_{\mathcal{B}}(t_1, t_2; \bar{\xi}_1, \bar{\xi}_2) d\omega_{\xi}}{K_{\mathcal{B}}(t_1, t_2; \bar{t}_1, \bar{t}_2)} K_{\mathcal{B}}(z_1, z_2; \bar{t}_1, \bar{t}_2). \end{aligned}$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,  
CAMBRIDGE, MASS.





## ANALYTIC SYSTEMS OF CENTRAL CONICS IN SPACE

BY  
J. L. COOLIDGE

The amount of literature dealing with conic sections, individual curves and systems of curves, in one plane, is vast. When however we are dealing with a number of conics, not in the same plane, the situation is quite different. Certain figures, as the focal conics of a set of confocal quadrics, are familiar enough, but very little has been done in the way of a systematic study of more general systems. There are some studies carried out with the aid of purely synthetic methods; the algebraic or analytic treatment lags behind.

The first writer to suggest a reasonable set of coordinates for a conic in space was Spottiswoode<sup>(1)</sup>. The totality of straight lines that intersect a conic in three-space generates a very special sort of quadratic complex. The coefficients determining the equation of this complex, when a straight line has the usual Plücker line coordinates, may be taken as the coordinates of the conic, a clumsy enough system. A much better technique, perhaps the best for algebraic purposes, was developed by Johnson<sup>(2)</sup>. Here a conic is looked upon not as a locus, but as the envelop of its tangent planes. Thus its tangential equation  $a^{ij}u_i u_j = 0$  gives ten homogeneous coordinates, connected by a quartic identity

$$a^{ij} = a^{ji}, \quad |a^{ij}| = 0, \quad i, j = 1, 2, 3, 4.$$

I think that this gives the best approach to the study of algebraic systems of conics, and I regret that more attention has not been given to the subject. For instance a complete study of linear and quadratic systems would be interesting. When it comes to attacking differential properties of conics this technique is disappointing, even as the Plücker line coordinates are of comparatively little use in studying the differential properties of systems of lines. In fact as far as I can make out very little has been written about the differential geometry of systems of conics. The most important article I have been able to find was by Blutel<sup>(3)</sup>, and his problem is very special. In what follows I am going to outline what seems to me the most promising way to approach the subject, and give a certain number of theorems. I hope that others may feel inclined to carry the study further, even though present mathematical fashion is concerned with very different questions.

Presented to the Society, April 27, 1940; received by the editors December 1, 1939.

<sup>(1)</sup> *On the twenty-one coordinates of a conic in space*, Transactions of the London Mathematical Society, vol. 10 (1879).

<sup>(2)</sup> *The conic as a space element*, these Transactions, vol. 15 (1914).

<sup>(3)</sup> *Recherches sur les surfaces qui sont en même temps lieux de coniques et enveloppes de cones*, Annales de l'École Normale Supérieure, (3), vol. 7 (1890).

1. **Series of conics.** The most obvious way to approach the study of the conic in space is to treat it as a rational curve. Let us use nonhomogeneous Cartesian coordinates, and assume that our conic is not a parabola. We may then write its parametric equations in the form

$$x^i = a^i t + b^i + c^i \frac{1}{t}, \quad i = 1, 2, 3.$$

Here, (a) and (c) give the directions of the asymptotes, (b) are the coordinates of the centre and, if we assume that  $t=1$  gives a vertex,  $t^2$  is the ratio of the distances from the asymptotes.

I give these equations because they seem to offer a favorable opening for the study of systems of conics, and in fact I personally first tried the problem in this way. I hasten to add that I was not at all able to attain the results which I believe to be easily attainable.

I turn to a different method which seems to fit the case even better. This is the method of moving axes first developed by Darboux in the opening chapters of his *Théorie Générale des Surfaces*, and extended in recent years by Cartan. Let a point have the rectangular Cartesian coordinates ( $X^i$ ) with regard to a set of fixed axes. Its coordinates with respect to a moving set of such axes shall be ( $x^i$ ). The coordinates with regard to the fixed axes of the moving origin shall be ( $X_0^i$ ). We then have the fundamental relations

$$(1) \quad X^i = X_0^i + a_{ij} x^j, \quad a_{ij} = \frac{\partial |a_{pq}|}{\partial a_{ij}}; \quad |a_{pq}| = 1.$$

Let the position of the point and also the situation of the moving axes be functions of a parameter  $v$ , which for simplicity of language I shall call "time." We then have

$$(2) \quad \frac{\partial X^i}{\partial v} = \frac{\partial X_0^i}{\partial v} + \frac{\partial a_{ij}}{\partial v} x^j + a_{ij} \frac{\delta x^j}{\delta v}.$$

I now seek the components with regard to the moving axes of the total velocity of the point. We write these  $\partial x^i / \partial v$ , while we mean by the notation  $\delta x^i / \delta v$  the velocity with regard to the moving axes of the point's motion with regard to those same axes:

$$(3) \quad \frac{\partial x^i}{\partial v} = \xi^i + \frac{\delta x^i}{\delta v} + q_{ij} x^j,$$

$$(4) \quad q_{ij} = -q_{ji} = a_{ki} \frac{\partial a_{kj}}{\partial v}, \quad \xi^i = a_{ki} \frac{\partial X_0^k}{\partial v}.$$

It is to be remembered that  $\|a_{ij}\|$  is the matrix of an orthogonal substitution of determinant 1.

These are the general formulae for moving rectangular axes in any number of dimensions. For our particular problem, let us assume that our conic lies in the plane  $x^3=0$  and that it is expressed parametrically

$$(5) \quad x^1 = a \cos u, \quad x^2 = b \sin u, \quad x^3 = 0.$$

Let us further simplify the notation by writing  $i, j, k$  as a cyclic permutation of 1, 2, 3 and putting

$$(6) \quad q_{ij} = -p_k.$$

We have then our fundamental formulae

$$(7) \quad \begin{aligned} \frac{\partial x^1}{\partial u} &= -a \sin u, & \frac{\partial x^1}{\partial v} &= \xi^1 + \frac{\partial a}{\partial v} \cos u - p_3 b \sin u, \\ \frac{\partial x^2}{\partial u} &= b \cos u, & \frac{\partial x^2}{\partial v} &= \xi^2 + \frac{\partial b}{\partial v} \sin u + p_3 a \cos u, \\ \frac{\partial x^3}{\partial u} &= 0, & \frac{\partial x^3}{\partial v} &= \xi^3 + p_1 b \sin u - p_2 a \cos u. \end{aligned}$$

We have further in the classical notation

$$(8) \quad \begin{aligned} E &= a^2 \sin^2 u + b^2 \cos^2 u, \\ F &= b\xi^2 \cos u - a\xi^1 \sin u + \frac{1}{2} \frac{\partial(b^2 - a^2)}{\partial v} \cos u \sin u + p_3 ab, \\ G &= \sum_i (\xi^i)^2 + 2 \left[ \xi^1 \frac{\partial a}{\partial v} + (\xi^2 p_3 - \xi^3 p_2) a \right] \cos u \\ &\quad + 2 \left[ \xi^2 \frac{\partial b}{\partial v} + (\xi^3 p_1 - \xi^1 p_3) b \right] \sin u \\ &\quad + \left[ (p_2^2 + p_3^2) a^2 + \left( \frac{\partial a}{\partial v} \right)^2 \right] \cos^2 u \\ &\quad - 2 \left[ p_3 \left( b \frac{\partial a}{\partial v} - a \frac{\partial b}{\partial v} \right) + p_1 p_2 ab \right] \cos u \sin u \\ &\quad + \left[ (p_3^2 + p_1^2) b^2 + \left( \frac{\partial b}{\partial v} \right)^2 \right] \sin^2 u, \end{aligned}$$

$$(9) \quad \begin{aligned} \frac{\partial(x^1, x^2)}{\partial(u, v)} &= - \left[ a\xi^2 \sin u + b\xi^1 \cos u + b \frac{\partial a}{\partial v} \cos^2 u \right. \\ &\quad \left. - p_3(b^2 - a^2) \cos u \sin u + a \frac{\partial b}{\partial v} \sin^2 u \right], \end{aligned}$$

$$(EG - F^2)^{1/2}D = -ab \frac{\partial x^3}{\partial v},$$

$$(10) (EG - F^2)^{1/2}D' = \frac{\partial^2 x^3}{\partial u \partial v} \frac{\partial(x^1, x^2)}{\partial(u, v)} - \frac{\partial x^3}{\partial v} \left[ \left( b \frac{\partial a}{\partial v} - a \frac{\partial b}{\partial v} \right) \cos u \sin u + p_3(b^2 \cos^2 u + a^2 \sin^2 u) \right].$$

If we have given two conics, they may have any one of the five following relations: (a) they do not intersect, (b) they intersect once, (c) they intersect twice, (d) they touch, and (e) they may be coplanar. Omitting the last case, when we are considering a one-parameter family of conics in space, we have to distinguish the cases where adjacent conics do not meet, or where they meet once, or where they meet twice, or where they are tangent to one another. Or to put the matter in more exact language, they may not all touch any curve, or they may all touch a curve, or they may touch two curves (or one curve twice), or they may all touch a curve and lie in the corresponding osculating planes. Let us verify these statements analytically.

If the conics of a series touch one curve, it must be possible to make  $u$  such a function of  $v$  that

$$\frac{\partial x^i}{\partial u} = \rho \frac{\partial x^i}{\partial v}.$$

This involves

$$\frac{\partial x^3}{\partial v} = 0, \quad \frac{\partial(x^1, x^2)}{\partial(u, v)} = 0.$$

If we replace the sine and cosine of  $u$  by  $(2t)/(1+t^2)$ ,  $(1-t^2)/(1+t^2)$ , where  $t$  is the tangent of half of the excentric angle, we have the condition that the resultant of a quadratic and a quartic polynomial in  $t$  should vanish, which is a bit long to write out, but involves no theoretical difficulties. There is more interest in the case where the conics touch two curves.

These curves lie on the developable surface generated by the plane of the conic, the characteristic line being the intersection with

$$\frac{\partial x^3}{\partial v} = \xi_3 + p_1 b \sin u - p_2 a \cos u = 0.$$

The points where  $\partial(x^1, x^2)/\partial(u, v) = 0$  must include the two intersections of the conic with  $\partial x^3/\partial v = 0$  so that

$$(11) \quad \frac{\partial(x^1, x^2)}{\partial(u, v)} = \frac{\partial x^3}{\partial v} (\alpha \cos u + \beta \sin u + \gamma).$$

Again look at the matter geometrically. When two conics intersect twice, they lie on a pencil of quadric surfaces, two of which are cones, and the vertices of these cones are harmonically separated by the planes of the conics. When the conics are infinitely near, one cone tends to be squashed between them. It appears then that if the conics of our series are twice tangent to a curve, the tangent planes to the surface generated at all points of a conic pass through a common point and envelop a cone. Now let us look at the matter analytically. The equation of the tangent plane is

$$(X^1 - x^1) \frac{\partial(x^2, x^3)}{\partial(u, v)} + (X^2 - x^2) \frac{\partial(x^3, x^1)}{\partial(u, v)} + (X^3 - x^3) \frac{\partial(x^1, x^2)}{\partial(u, v)} = 0,$$

$$[X^1(b \cos u) + X^2(a \sin u) - ab] \frac{\partial x^3}{\partial v} + X^3 \frac{\partial(x^1, x^2)}{\partial(u, v)} = 0.$$

The reader should not confuse  $X$  appearing here with that in (1). This becomes in the present instance, thanks to (11),

$$X^1 b \cos u + X^2 a \sin u + X^3 [\alpha \cos u + \beta \sin u + \gamma] - ab = 0.$$

It appears then that the point

$$X^1 = -a\alpha/\gamma, X^2 = -b\beta/\gamma, X^3 = ab/\gamma$$

is in the tangent plane at every point of the conic. Conversely, when these tangent planes pass through such a point, we have an identity in  $v$  and  $\partial(x^1, x^2)/\partial(u, v)$  is divisible by  $\partial x^3/\partial v$  so that the conics touch two curves or are the limits of conics touching two curves.

**THEOREM 1.** *If the conics of a series are tangent to two curves, the tangent planes to the surface generated at all points of a conic will envelop a quadric cone which touches the surface all along the conic.*

Now consider the dual. We have a one-parameter family of quadric cones. If adjacent cones tend to touch twice, that is to say, if the cones are inscribed in two developable surfaces, they will also intersect in a conic.

**THEOREM 2.** *If the quadric cones of a one-parameter series be inscribed in two different developables, the characteristic curves of these cones will be the generators of these developables and a series of conics tangent to two curves, or the limit of such a series.*

The surfaces generated by these conics are the ones considered by Blutel (q. v.).

There remains the case where adjacent conics tend to touch. This means that  $\partial(x^1, x^2)/\partial(u, v)$  is divisible by  $\partial x^3/\partial v$  but the line

$$\frac{\partial x^3}{\partial v} = 0$$



is tangent to the conic. If  $P$  be the point of contact, its line of advance is along the conic, and also along the characteristic line whose equation has just been written. Hence  $P$  must be the point of contact with the edge of regression. The plane of the conic must then be the osculating plane for the curve generated by  $P$ . Hence we have a series of conics tangent to a curve each lying in the corresponding osculating plane. Here also there will be a quadric cone tangent at all points of the conic.

We have assumed (11) here, that is,

$$\frac{\partial(x^1, x^2)}{\partial(u, v)} = \frac{\partial x^3}{\partial v} (\alpha \cos u + \beta \sin u + \gamma).$$

This identity will lead to the equations

$$(12) \quad \begin{aligned} -p_2 a \alpha + \gamma \xi^3 &= -b \frac{\partial a}{\partial v}, & p_1 b \beta + \gamma \xi^3 &= -a \frac{\partial b}{\partial v}, \\ p_1 b \alpha - p_2 a \beta &= p_3 (a^2 - b^2), \\ \alpha \xi^3 - \gamma a p_2 &= -b \xi^1, & \beta \xi^3 + \gamma b p_1 &= -a \xi^2. \end{aligned}$$

We see geometrically that if two conics lie on the same quadric cone, the generators of the cone establish a projective relation between them. When the conics are infinitely near, the generators give the directions of the curves conjugate to the conics in the surface generated. Analytically, the differential equation for the curves conjugate to the conics  $\delta v = 0$  is

$$\begin{aligned} Ddu + D'dv &= 0, \\ -ab \frac{\partial x^3}{\partial v} du + \left[ \frac{\partial^2 x^3}{\partial u \partial v} \frac{\partial(x^1, x^2)}{\partial(u, v)} - \frac{\partial x^3}{\partial v} \left\{ b \frac{\partial a}{\partial v} - a \frac{\partial b}{\partial v} \right\} \cos u \sin u \right. \\ &\quad \left. + p_3 (b^2 \cos^2 u + a^2 \sin^2 u) \right] dv = 0. \end{aligned}$$

In the present case this is

$$\begin{aligned} -abdu + \left[ (p_1 b \cos u + p_2 a \sin u)(\alpha \cos u + \beta \sin u + \gamma) \right. \\ \left. - \left\{ \left( b \frac{\partial a}{\partial v} - a \frac{\partial b}{\partial v} \right) \cos u \sin u \right. \right. \\ \left. \left. + p_3 (b^2 \cos^2 u + a^2 \sin^2 u) \right\} \right] dv = 0. \end{aligned}$$

In view of (12) this becomes

$$du + [L(v) + M(v) \cos u + N(v) \sin u] dv = 0.$$

Let us now introduce the tangent of the half-angle, so that

$$\cos u = \frac{1-t^2}{1+t^2}, \quad \sin u = \frac{2t}{1+t^2}, \quad du = \frac{dt}{1+t^2},$$

$$\frac{dt}{dv} + A(v) + B(v)t + C(v)t^2 = 0.$$

This is a Riccati equation, characterized by the fact that the cross ratio of four solutions is constant. This gives

**BLUTEL'S THEOREM 3.** *If the central conics of a series be not coplanar, but touch two curves, the conjugate curves on the surface they generate will establish a projective correspondence among them<sup>(4)</sup>.*

Let us now look at the orthogonal trajectories of conics. Their differential equation is

$$(13) \quad Edu + Fdv = 0.$$

Introducing the tangent of the half-angle as before, we have

$$2[b^2t^4 + (4a^2 - 2b^2)t^2 + b^2]dt + \left[ b\xi^2(1-t^4) - 2a\xi^1t(1+t^2) + \frac{\partial(b^2 - a^2)}{\partial v}t(1-t^2) + p_3ab(1+t^2)^2 \right](1+t^2)dv = 0.$$

These trajectories will establish a projective correspondence if this is a Riccati equation. If  $b^2 = a^2$ , the equation reduces automatically to the Riccati form. Suppose that  $b^2 \neq a^2$ . Then  $1+t^2$  cannot divide the coefficient of  $dt$  and the first factor in the coefficient of  $dv$  must be proportional to the coefficient of  $dt$ . Evidently, the factor of proportionality must be zero, so that  $F=0$  or

$$\xi^1 = 0, \quad \xi^2 = 0, \quad p_3 = 0, \quad b^2 - a^2 = k,$$

where  $k$  is a constant, not zero.

**THEOREM 4.** *The necessary and sufficient condition that the trajectories orthogonal to the central conics of a series should establish a projective correspondence among them is that the conics should be circles; or the centre should be fixed or move orthogonally to the plane, the distance between the foci should be constant and the axes should not twist.*

Let us now try to discover under what circumstances these orthogonal trajectories are geodesic curves of the surface. The necessary and sufficient condition for this is that

$$\frac{\partial}{\partial u} \left( \frac{EG - F^2}{E} \right) = 0.$$

<sup>(4)</sup> Blutel, loc. cit., p. 155.

This shows that  $F^2$ , and so  $F$ , is divisible by  $E$  when the roots of  $E$  are distinct, and as they are of the same order in  $t$  when we substitute the tangent of the half-angle, the factor must be a function of  $v$ :

$$F = f(v)E,$$

$$f(v)[a^2 \sin^2 u + b^2 \cos^2 u] \\ = \left[ b\xi^2 \cos u - a\xi^1 \sin u + \frac{1}{2} \frac{\partial(b^2 - a^2)}{\partial v} \cos u \sin u + p_3 ab \right].$$

It follows from this identity that  $f(v)=0$ ,  $F=0$  or  $\xi^1=\xi^2=p_3=0$ ,  $b^2-a^2=c$ , where  $c$  is a constant, not zero.

Thus  $\partial G/\partial u=0$ , or

$$\xi^3 p_1 b \cos u + \xi^2 p_2 a \sin u + p_1 p_2 ab (\sin^2 u - \cos^2 u) \\ + \left[ p_1^2 b^2 + \left( \frac{\partial b}{\partial v} \right)^2 - p_2^2 a^2 - \left( \frac{\partial a}{\partial v} \right)^2 \right] \sin u \cos u = 0.$$

Hence

$$\xi^3 p_1 = \xi^2 p_2 = p_1 p_2 = p_1^2 b^2 - p_2^2 a^2 + \left( \frac{\partial b}{\partial v} \right)^2 - \left( \frac{\partial a}{\partial v} \right)^2 = 0.$$

If  $\xi^3 \neq 0$ , then  $p_1=0$ ,  $p_2=0$ ,  $\partial a/\partial v = \partial b/\partial v = 0$ . We have a conic of fixed axes generating a right cylinder.

If  $\xi^3=0$ , we have a fixed centre, and either

$$p_1 = 0, \quad \left( \frac{\partial b}{\partial v} \right)^2 - \left( \frac{\partial a}{\partial v} \right)^2 = p_2^2 a^2,$$

or

$$p_2 = 0, \quad \left( \frac{\partial a}{\partial v} \right)^2 - \left( \frac{\partial b}{\partial v} \right)^2 = p_1^2 b^2.$$

The distance between the foci is constant, the plane rotates about one axis which has a fixed direction.

There is the second case where

$$E = a^2 = b^2, \quad F = [a\xi^2 \cos u - a\xi^1 \sin u + p_3 a^2].$$

We get from  $\partial(EG - F^2)/\partial u = 0$  that

$$G = [\xi^2 \cos u - \xi^1 \sin u + p_3 a]^2 + \phi(v), \\ \xi^1 \frac{\partial a}{\partial v} - \xi^2 a p_2 = 0, \quad \xi^2 \frac{\partial a}{\partial v} + \xi^3 a p_1 = 0, \\ a^2 (p_1^2 - p_2^2) = (\xi^1)^2 - (\xi^2)^2, \quad a^2 p_1 p_2 = \xi^1 \xi^2.$$

One solution of these equations is  $p_1 = p_2 = 0$ ,  $\xi^1 = \xi^2 = 0$ , and this gives rise to the parallels of a surface of revolution.

A second solution is  $\partial a / \partial v = 0$ ,  $\xi^3 = 0$ ,  $\xi^1 = \pm a p_1$ ,  $\xi^2 = \pm a p_2$ . It is readily shown that the last three of these equations constitute necessary and sufficient conditions that the centre of the moving circle lie on and move in the direction of the characteristic line of the plane of the circle and have a velocity which is  $\pm a$  times the angular velocity with which this plane turns about the characteristic line. The conditions also guarantee that the planes of the circles are the osculating planes of the locus of the centres. Thus, the circles have their centres in the points of a twisted curve of constant torsion  $1/a$ , lie in the osculating planes of this curve, and have the constant radius  $|a|^{(6)}$ .

All other solutions of the four equations are imaginary.

**THEOREM 5.** *If the central conics of a series are geodesically parallel, but are not circles, either they are the right sections of a quadric cylinder or the centre and direction of one axis is fixed, and the distance between the foci is constant. If the conics are real circles, either they are the parallels of a surface of revolution, or they have their centres in the points of a twisted curve of constant torsion  $1/a$ , lie in the osculating planes of this curve and have the constant radius  $|a|^{(6)}$ .*

Let us next inquire under what circumstances the conics will be lines of curvature. A plane curve will be a line of curvature if the normals to the surface all along it make the same angle with the plane, or what comes to the same thing, the normals to the curve making a certain constant angle with the plane are normal to the surface. If  $C$  be the tangent of the angle which a normal to the conic makes with the  $x^3$  axis, the direction cosines of this normal are proportional to

$$b \cos u, \quad a \sin u, \quad -C(b^2 \cos^2 u + a^2 \sin^2 u)^{1/2}.$$

This will be normal to the surface, that is to say, normal to  $\partial x / \partial v$  if

$$\left[ \xi^1 b \cos u + \xi^2 a \sin u + b \frac{\partial a}{\partial v} \cos^2 u + a \frac{\partial b}{\partial v} \sin^2 u - p_3(b^2 - a^2) \cos u \sin u \right]^2 \\ = C^2(b^2 \cos^2 u + a^2 \sin^2 u)(\xi^3 + p_1 b \sin u - p_2 a \cos u)^2.$$

This is to be an identity in  $u$ . The left side is a perfect square, hence either the right side is, or both vanish identically. Excluding this case, the right is a perfect square if  $b^2 \cos^2 u + a^2 \sin^2 u$  is a perfect square, and this involves  $a = b$  so that we have a circle. The evolutes of a circle are points, a sphere will touch the surface all along the circle, or the surface is the envelop of a one-parameter family of spheres.

Suppose, next, that each side vanishes identically and that  $C = 0$ . Then

$$\xi^1 = \xi^2 = \frac{\partial a}{\partial v} = \frac{\partial b}{\partial v} = p_3(b^2 - a^2) = 0.$$

<sup>(6)</sup> This possibility was pointed out to me by Professor Graustein.

$C=0$  gives the fact that the plane of the conic is orthogonal to the surface,  $\xi^1=\xi^2=0$  the centre is fixed, or moves orthogonally to the plane,  $\partial a/\partial v=\partial b/\partial v=0$  that the lengths of the axes are constant. If  $p_3=0$ , then  $\partial x^1/\partial v=\partial x^2/\partial v=0$  and every point moves orthogonally to the plane. If  $b=a$ , the surface is the envelop of spheres of constant radius, and is therefore a canal surface.

If, on the other hand,  $C\neq 0$ , then  $\xi^3=p_1=p_2=0$  and  $\partial x^i/\partial v=0$ , so that there is no surface, unless  $b=a$ , in which case we have the circles as before.

**THEOREM 6.** *The necessary and sufficient condition that the central conics of a non-planar series should be lines of curvature is that either they be the characteristic circles on a one-parameter family of spheres, or that they be invariable in size and shape and invariant in their planes and so generate a surface of Monge.*

**2. Congruences.** Let us pass to two-parameter systems, or congruences. Let us call the parameters  $v_1, v_2$ , putting subscripts 1 or 2 to the notations of (3), (4), (6), (7), (8) to indicate the variable with regard to which the differentiation has been performed. Let us look for the focal points, which we described geometrically as the points where a conic meets an infinitely near one. Analytically this means that when a certain relation has been established between  $v_1$  and  $v_2$  we can make  $u$  such a function of these variables that the tangent to the curve traced is the same as that to the conic. This again will involve three relations

$$\frac{\partial x^i}{\partial v_1} dv_1 + \frac{\partial x^i}{\partial v_2} dv_2 + \lambda \frac{\partial x^i}{\partial u} du = 0.$$

Setting the discriminant of these three linear homogeneous equations equal to 0, we get a cubic expression in  $\cos u, \sin u$  which will have six roots.

**THEOREM 7.** *The central conics of a congruence will usually have six focal points where they touch six surfaces or meet certain curves.*

This number is in accordance with a result of Darboux's where it is shown<sup>(\*)</sup> that where a congruence is composed of plane curves of order  $m$  the number of focal points is  $m(m+1)$ . When our central conics are circles, two focal points are on the circle at infinity, we usually overlook them and say that the circles of a congruence touch four surfaces.

Let us now inquire under what circumstances the conics of a congruence are orthogonal to a surface. For this purpose,  $u$  must be such a function of  $v_1$  and  $v_2$  that

$$E \frac{\partial u}{\partial v_1} + F_1 = E \frac{\partial u}{\partial v_2} + F_2 = 0.$$

(\*) *Théorie Générale des Surfaces*, vol. 2, p. 4.



The condition of integrability will be

$$(14) \quad E \left( \frac{\partial F_1}{\partial v_2} - \frac{\partial F_2}{\partial v_1} \right) + F_1 \left( \frac{\partial F_2}{\partial u} - \frac{\partial E}{\partial v_2} \right) + F_2 \left( \frac{\partial E}{\partial v_1} - \frac{\partial F_1}{\partial u} \right) = 0.$$

Developing this at length, we get a rather fearsome equation

$$(15) \quad \begin{aligned} & \frac{ab}{2} \left[ p_{22} \frac{\partial(a^2 + b^2)}{\partial v_1} - p_{31} \frac{\partial(a^2 + b^2)}{\partial v_2} + 2(\xi_1^1 \xi_2^2 - \xi_2^1 \xi_1^2) \right] \\ & + \cos u \left[ \frac{b \xi_2^2}{2} \frac{\partial(a^2 + b^2)}{\partial v_1} - \frac{b \xi_1^2}{2} \frac{\partial(a^2 + b^2)}{\partial v_2} + a^2 b (p_{21} \xi_1^1 - p_{31} \xi_2^1) \right] \\ & + \sin u \left[ \frac{a \xi_1^1}{2} \frac{\partial(a^2 + b^2)}{\partial v_2} - \frac{a \xi_2^1}{2} \frac{\partial(a^2 + b^2)}{\partial v_1} + ab^2 (p_{21} \xi_1^2 - p_{31} \xi_2^2) \right] \\ & + (b^2 \cos^2 u + a^2 \sin^2 u) \left( \frac{\partial(ab p_{22})}{\partial v_1} - \frac{\partial(ab p_{31})}{\partial v_2} \right) \\ & + \frac{\sin u \cos u}{4} \left[ \frac{\partial((b^2 + a^2), (b^2 - a^2))}{\partial(v_1, v_2)} \right] + \cos^3 u b^2 \left( \frac{\partial b \xi_1^2}{\partial v_2} - \frac{\partial b \xi_2^2}{\partial v_1} \right) \\ & + \cos^2 u \sin u \left[ b^2 \left( \frac{\partial a \xi_2^1}{\partial v_1} - \frac{\partial a \xi_1^1}{\partial v_2} \right) + \frac{a}{2} \left( \frac{\partial(b^2 - a^2)}{\partial v_2} \xi_1^1 - \frac{\partial(b^2 - a^2)}{\partial v_1} \xi_2^1 \right) \right] \\ & + \cos u \sin^2 u \left[ a^2 \left( \frac{\partial b \xi_1^2}{\partial v_2} - \frac{\partial b \xi_2^2}{\partial v_1} \right) + \frac{b}{2} \left( \frac{\partial(b^2 - a^2)}{\partial v_2} \xi_1^2 - \frac{\partial(b^2 - a^2)}{\partial v_1} \xi_2^2 \right) \right] \\ & + \sin^3 u a^2 \left( \frac{\partial a \xi_2^1}{\partial v_1} - \frac{\partial a \xi_1^1}{\partial v_2} \right) = 0. \end{aligned}$$

If we replace  $u$  by  $t$  as before, this equation becomes sextic.

**THEOREM 8.** *If the central conics of a congruence be normal to more than six surfaces, the congruence is a normal one.*

The conditions for a normal congruence will, then, be that the left side of this equation should vanish identically, for all values of  $u$ . Now if we have an expression

$$A_0 + A_1 \cos u + B_1 \sin u + A_2 \cos^2 u + B_2 \cos u \sin u + C_2 \sin^2 u + A_3 \cos^3 u + B_3 \cos^2 u \sin u + C_3 \cos u \sin^2 u + D_3 \sin^3 u = 0,$$

we shall find the conditions for vanishing identically by putting  $u$  successively equal to  $0, \pi/4, \pi, 2, 3\pi/4, \pi, 3\pi/4, 3\pi/2, 2\pi/4$ . This will give

$$B_2 = A_0 + A_2 = A_0 + C_2 = A_1 + A_3 = B_1 + D_3 = A_1 + C_3 = B_1 + B_3 = 0.$$

Assuming, then, that  $b^2 \neq a^2$ , we find

$$\begin{aligned}
 A_0 = A_2 = B_2 = C_2 = 0; A_1 + A_3 = A_1 + C_3 = B_1 + B_3 = B_1 + D_3 = 0; \\
 \text{(I)} \quad \frac{\partial(a, b)}{\partial(v_1, v_2)} = 0; \\
 \text{(II)} \quad \frac{\partial(p_{31}ab)}{\partial v_2} - \frac{\partial(p_{32}ab)}{\partial v_1} = 0; \\
 \text{(III)} \quad \frac{1}{2} \left[ p_{32} \frac{\partial(a^2 + b^2)}{\partial v_1} - p_{31} \frac{\partial(a^2 + b^2)}{\partial v_2} \right] + \xi_1^1 \xi_2^2 - \xi_1^2 \xi_2^1 = 0; \\
 \text{(IV)} \quad \frac{\partial}{\partial v_2} \left( \frac{b \xi_1^2}{(b^2 - a^2)^{1/2}} \right) = \frac{\partial}{\partial v_1} \left( \frac{b \xi_2^2}{(b^2 - a^2)^{1/2}} \right); \\
 \text{(V)} \quad \frac{\partial}{\partial v_2} \left( \frac{a \xi_1^1}{(b^2 - a^2)^{1/2}} \right) = \frac{\partial}{\partial v_1} \left( \frac{a \xi_2^1}{(b^2 - a^2)^{1/2}} \right); \\
 \text{(VI)} \quad (p_{32} \xi_1^1 - p_{31} \xi_2^1) + \xi_2^2 \frac{\partial}{\partial v_1} \log \frac{a^2}{(b^2 - a^2)^{1/2}} - \xi_1^2 \frac{\partial}{\partial v_2} \log \frac{a^2}{(b^2 - a^2)^{1/2}} = 0; \\
 \text{(VII)} \quad (p_{32} \xi_1^2 - p_{31} \xi_2^2) - \xi_2^1 \frac{\partial}{\partial v_1} \log \frac{b^2}{(b^2 - a^2)^{1/2}} + \xi_1^1 \frac{\partial}{\partial v_2} \log \frac{b^2}{(b^2 - a^2)^{1/2}} = 0.
 \end{aligned}$$

The most important of these equations is (I) which gives

**THEOREM 9.** *The semi-axes of the central conics of a normal congruence are functionally related.*

Let us next assume that we are in the special case where

$$(16) \quad \xi_1^1 \xi_2^2 - \xi_1^2 \xi_2^1 = 0.$$

This means geometrically that either we have a fixed centre, or the centre traces a curve, or that at each centre the plane of the conic is orthogonal to the tangent plane to the surface of centres. From (III) follows

**THEOREM 10.** *If the central conics of a normal congruence have axes of fixed lengths, the centre will be fixed, or trace a curve, or a surface orthogonal at each point to the plane of the corresponding conic.*

Assuming that (16) still holds, but the lengths of both axes are not fixed, we write

$$\begin{aligned}
 \xi^1 dt &= \xi_1^1 \frac{dv_1}{dt} dt + \xi_2^1 \frac{dv_2}{dt} dt = 0, \\
 \xi^2 dt &= \xi_1^2 \frac{dv_1}{dt} dt + \xi_2^2 \frac{dv_2}{dt} dt = 0,
 \end{aligned}
 \quad (17)$$

$$(18) \quad \begin{aligned} da &= \frac{\partial a}{\partial v_1} \frac{dv_1}{dt} dt + \frac{\partial a}{\partial v_2} \frac{dv_2}{dt} dt = 0, \\ p_3 dt &= p_{31} \frac{dv_1}{dt} dt + p_{32} \frac{dv_2}{dt} dt = 0. \end{aligned}$$

Here  $t$  is an arbitrary variable not  $\tan v/2$ . By (16) the first two have a common solution, and then, by (I), (II) and (III) the last two have a common solution unless  $a^2 + b^2 = \text{const.}$  Suppose, first, that all four have a common solution. We are then at liberty to assume

$$(19) \quad \frac{\partial a}{\partial v_2} = \frac{\partial b}{\partial v_2} = p_{32} = \xi_2^1 = \xi_2^2 = 0.$$

We have also the additional equations

$$(20) \quad \frac{\partial \xi_1^1}{\partial v_2} = \frac{\partial \xi_1^2}{\partial v_2} = \frac{\partial p_{31}}{\partial v_2} = 0.$$

All seven of our equations (I)-(VII) are satisfied. Assuming that we have a surface of centres, the curves  $v_1 = \text{const.}$  thereon correspond to constant values for the lengths of the axes. Let us take  $v_2 = \text{const.}$  as the curves orthogonal to them. The equations  $\xi_2^1 = \xi_2^2 = 0$  tell us that the curves  $v_1 = \text{const.}$  are orthogonal to the corresponding planes of the conics, hence the curves  $v_2 = \text{const.}$  are tangent to the planes of the conics, or lie in them. But now we find from the first two equations (7) that the instantaneous motion of every point of the conic, when  $v_1$  is constant, is orthogonal to the plane. The equations

$$\frac{\partial \xi_1^1}{\partial v_2} = \frac{\partial \xi_1^2}{\partial v_2} = 0$$

give by (4)

$$\frac{\partial}{\partial v_2} \left( a_{k1} \frac{\partial X_0^k}{\partial v_1} \right) = \frac{\partial}{\partial v_2} \left( a_{k2} \frac{\partial X_0^k}{\partial v_1} \right) = 0.$$

Since  $\partial X_0 / \partial v_1$  is in the plane of the conic

$$\begin{aligned} \frac{\partial X_0^i}{\partial v_1} &= \lambda a_{j1} + \mu a_{j2} = \left( a_{k1} \frac{\partial X_0^k}{\partial v_1} \right) a_{j1} + \left( a_{k2} \frac{\partial X_0^k}{\partial v_1} \right) a_{j2}, \\ \frac{\partial^2 X_0^i}{\partial v_1 \partial v_2} &= \left( a_{k1} \frac{\partial X_0^k}{\partial v_1} \right) \frac{\partial a_{j1}}{\partial v_2} + \left( a_{k2} \frac{\partial X_0^k}{\partial v_1} \right) \frac{\partial a_{j2}}{\partial v_2}. \end{aligned}$$

But by (4) and  $p_{32} = 0$

$$\left(a_{j1} \frac{\partial a_{j1}}{\partial v_2}\right) = \left(a_{j1} \frac{\partial a_{j2}}{\partial v_2}\right) = \left(a_{j2} \frac{\partial a_{j1}}{\partial v_2}\right) = \left(a_{j2} \frac{\partial a_{j2}}{\partial v_2}\right) = 0.$$

Hence

$$\left(a_{j1} \frac{\partial^2 X_0^j}{\partial v_1 \partial v_2}\right) = \left(a_{j2} \frac{\partial^2 X_0^j}{\partial v_1 \partial v_2}\right) = 0.$$

Hence  $\partial^2 X_0^j / \partial v_1 \partial v_2 = \rho (\partial X_0^j / \partial v_2)$  and

$$\frac{\partial}{\partial v_1} \left( \frac{\frac{\partial X_0^1}{\partial v_2}}{\frac{\partial X_0^3}{\partial v_2}} \right) = \frac{\partial}{\partial v_1} \left( \frac{\frac{\partial X_0^2}{\partial v_2}}{\frac{\partial X_0^1}{\partial v_2}} \right) = 0.$$

This means that the normals to the plane all along the curve  $v_2 = \text{const.}$  are parallel and this curve must be in the plane, not merely tangent to it. There are, hence, only a singly infinite number of planes, so that in each there are an infinite number of conics. We have moreover the equations

$$\frac{\partial \xi_1^1}{\partial v_2} = \frac{\partial \xi_1^2}{\partial v_2} = \frac{\partial p_{31}}{\partial v_2} = 0.$$

From the equations (7) when  $v = v_1$  everything is independent of  $v_2$ . Hence we have the same series of conics in all our planes. The congruence is generated by an immovable set of conics in a singly infinite set of planes. Conversely, such a set of conics will clearly generate a normal congruence.

When the centre traces a curve, if  $\xi_2^3 \neq 0$  we may repeat our previous reasoning,  $\partial X_0^1 / \partial v_2 \neq 0$  and we have in each plane a one-parameter family of concentric conics. If  $\xi_2^3 = 0$ ,  $\partial X_0^1 / \partial v_2 = 0$ , then

$$\frac{\partial}{\partial v_2} \left( a_{k1} \frac{\partial X_0^k}{\partial v_1} \right) = \frac{\partial}{\partial v_2} \left( a_{k2} \frac{\partial X_0^k}{\partial v_1} \right) = 0, \quad \frac{\partial a_{k1}}{\partial v_2} \frac{\partial X_0^k}{\partial v_1} = \frac{\partial a_{k2}}{\partial v_2} \frac{\partial X_0^k}{\partial v_1} = 0.$$

As before  $\partial X_0^j / \partial v_1 = \lambda a_{j1} + \mu a_{j2}$ ,  $\xi_1^3 = 0$ . This means that the curve traced by the centre is tangent to the plane of the conic. Our last expression can be written better

$$\frac{\partial X_0^j}{\partial v_1} = \left( a_{k1} \frac{\partial X_0^k}{\partial v_1} \right) a_{j1} + \left( a_{k2} \frac{\partial X_0^k}{\partial v_1} \right) a_{j2}.$$

Remembering that the left is independent of  $v_2$  and  $p_{31} = 0$ ,

$$\frac{\partial X_0^j}{\partial v_1} \frac{\partial a_{j1}}{\partial v_1} = \left( a_{k1} \frac{\partial X_0^k}{\partial v_1} \right) p_{31}.$$

Differentiating to  $v_2$ , remembering  $\partial X_0^j / \partial v_2 = p_{3j} = \partial p_{3j} / \partial v_2 = 0$ ,

$$\frac{\partial X_0^j}{\partial v_1} \frac{\partial^2 a_{j1}}{\partial v_1 \partial v_2} = \frac{\partial X_0^j}{\partial v_1} \frac{\partial^2 a_{j2}}{\partial v_1 \partial v_2} = 0.$$

Differentiating to  $v_1$ ,

$$\frac{\partial^2 X_0^j}{\partial v_1^2} \frac{\partial a_{j2}}{\partial v_2} = \frac{\partial^2 X_0^j}{\partial v_1^2} \frac{\partial a_{j1}}{\partial v_2} = 0.$$

Hence

$$A \frac{\partial^2 X_0^j}{\partial v_1^2} + B \frac{\partial X_0^j}{\partial v_1} + C X_0^j = 0.$$

The centre traces a plane curve through the fixed origin. As this can be anywhere, the curve must be a straight line. Hence we have a series of conics whose centres lie on a line while the plane is rotated about that line.

There remains the possibility that the centre of the conics should be fixed. Here we are back on the first case, we have a set of invariable conics whose planes envelop a cone with its vertex at their common centre. The reasoning is reversible in each case. We note that in every case we have a one-parameter family of conics immovable in a moving plane.

I return to the equations (17) and (18) and assume that the first two still have a common solution, and hence, that the last two have, when  $a^2 + b^2$  is not a constant, but that the solutions are different. Here we are free to assume that

$$\frac{\partial a}{\partial v_2} = p_{32} = 0, \quad \xi_1^1 = \xi_1^2 = 0.$$

From (IV) and (V),

$$\xi_2^1 = \frac{(b^2 - a^2)^{1/2}}{a} \phi_2(v_2), \quad \xi_2^2 = \frac{(b^2 - a^2)^{1/2}}{b} \psi_2(v_2).$$

From (II),  $p_{31}$  is a function of  $v_1$  alone. From (VI),  $\xi_2^2 / \xi_2^1$  is a function of  $v_1$  alone. Hence  $\psi_2 / \phi_2$  is a constant. We may change variables writing  $\phi_2 = v_2$ ,  $\psi_2 = kv_2$ . Then, from (VI) and (VII)

$$p_{31} = \frac{a}{kb} \frac{\partial}{\partial v_1} \log \frac{a^2}{(b^2 - a^2)^{1/2}} = - \frac{kb}{a} \frac{\partial}{\partial v_1} \log \frac{b^2}{(b^2 - a^2)^{1/2}},$$

$$p_{31} \left( k \frac{b}{a} + \frac{a}{kb} \right) = 2 \frac{\partial}{\partial v_1} \log \frac{a}{b}.$$



Putting  $a/b = \rho$ ,

$$p_{s_1} = \frac{2(k^2\rho^2 + 1)}{k\rho^2} \frac{\partial \rho}{\partial v_1}.$$

Again, eliminating  $p_{s_1}$  from (VI) and (VII)

$$a^2 \frac{\partial}{\partial v_1} \log \frac{a^2}{(b^2 - a^2)^{1/2}} + k^2 b^2 \frac{\partial}{\partial v_1} \log \frac{b^2}{(b^2 - a^2)^{1/2}} = 0.$$

This can be written

$$2a \frac{\partial a}{\partial v_1} + 2k^2 b^2 \frac{\partial b}{\partial v_1} = (a^2 + k^2 b^2) \frac{\partial \log (b^2 - a^2)^{1/2}}{\partial v_1},$$

$$a^2 + k^2 b^2 = C(b^2 - a^2),$$

where  $C$  is a constant since  $a$  and  $b$  are not functions of  $v_2$ . Hence  $\rho = \text{const.}$ ,  $p_{s_1} = 0$ . Hence from (VI) and (VII) we either have  $a$  and  $b$  constant which gives  $\partial a / \partial v_1 = \partial b / \partial v_1 = 0$  and throws us back on a previous case, or else

$$\xi_2^2 = \xi_2^1 = p_{s_2} = 0.$$

Here the centre is either fixed or traces a curve orthogonal to the plane of the conic. Again we are on a previous case.

Next we assume  $a^2 + b^2 = \text{const.}$  Then from (III) the equations (17) have a common solution. Assume that this is a solution of the first equation (18). We may write

$$\xi_2^1 = \xi_2^2 = \frac{\partial a}{\partial v_2} = 0.$$

From (VI) and (VII)

$$p_{s_2} \xi_1^1 = p_{s_2} \xi_1^2 = 0.$$

If  $p_{s_2} = 0$ , we have equations (19) and (20) and we can proceed as before. If  $\xi_1^1 = \xi_1^2 = 0$ , the centre is either fixed or traces a curve which cuts the plane orthogonally. There is but a singly infinite set of planes. We may take the parameter  $v_2$  to give the plane. As  $\partial a / \partial v_2 = \partial b / \partial v_2 = 0$ , we have in each plane the same set of concentric conics.

Suppose now that the solutions of (17) do not give a solution of (18).

We may write

$$\frac{\partial a}{\partial v_2} = \frac{\partial b}{\partial v_2} = 0, \quad \xi_1^1 = \xi_1^2 = 0.$$

We may assume  $a = \cos v_1$ ,  $b = \sin v_1$ . From (V) and (IV)

$$\xi_2^1 = \phi_1(v_2)(\tan^2 v_1 - 1)^{1/2}, \quad \xi_2^2 = \phi_2(v_2) \frac{(\tan^2 v_1 - 1)^{1/2}}{\tan v_1}.$$

Now if  $p_{31}=0$ ,  $\partial p_{31}/\partial v_1=0$ , and we are back on a previous case. If  $p_{31}\neq 0$ , we eliminate it between the equations (VI) and (VII) and find that  $\xi_2^2/\xi_1^2$  is a function of  $v_1$  alone. Hence  $\phi_2(v_2)=k(\phi_1(v_1))$  where  $k$  is constant, and

$$\begin{aligned}\partial \log \frac{a^2}{(b^2 - a^2)^{1/2}} + \left(\frac{\xi_2^1}{\xi_1^1}\right)^2 \partial \log \frac{b^2}{(b^2 - a^2)^{1/2}} &= 0, \\ \frac{-\tan^2 v_1}{\tan^2 v_1 - 1} - \frac{\operatorname{ctn}^2 v_1 \tan^2 v_1}{1 - \operatorname{ctn}^2 v_1} k^2 &= 0, \\ \frac{\tan^2 v_1 + k^2}{\tan^2 v_1 - 1} &= 0,\end{aligned}$$

so that  $\tan v_1$  is constant and we do not have a two-parameter family.

**THEOREM 11.** *If in a normal congruence of central conics the locus of the centres is a surface which at each point is orthogonal to the plane of the corresponding conic, or is a curve, or is fixed, the congruence is generated by a series of conics which are immovable in a moving plane.*

There remains the case where the first two equations (17) are not proportional. We may assume here  $\xi_3^1=\xi_1^2=0$ .

Let us write

$$\log \frac{a^2}{(b^2 - a^2)^{1/2}} = A, \quad \log \frac{b^2}{(b^2 - a^2)^{1/2}} = B.$$

From (IV) and (V)

$$\xi_2^2 = \frac{(b^2 - a^2)^{1/2}}{b} \phi_2(v_2), \quad \xi_1^1 = \frac{(b^2 - a^2)^{1/2}}{a} \phi_1(v_1).$$

We now change the variables  $v_1$  and  $v_2$  to such functions of them that

$$\begin{aligned}\frac{\partial w_1}{\partial v_1} &= \phi_1, & \frac{\partial w_2}{\partial v_2} &= \phi_2, \\ \phi_1(v_1) &= \psi_1(w_1), & \phi_2(v_2) &= \psi_2(w_2), \\ p_{32} &= \frac{-a\psi_2}{b} \frac{\partial A}{\partial w_1}, & p_{31} &= \frac{b\psi_1}{a} \frac{\partial B}{\partial w_2}.\end{aligned}$$

Further, let

$$a^2 \frac{\partial A}{\partial v_1} = \alpha, \quad b^2 \frac{\partial B}{\partial w_2} = \beta.$$

From (II) and (III)

$$(21) \quad \frac{\partial \alpha}{\partial w_1} + \frac{\partial \beta}{\partial w_2} = 0,$$

$$(22) \quad \alpha \frac{\partial(a^2 + b^2)}{\partial w_1} + \beta \frac{\partial(a^2 + b^2)}{\partial w_2} = 0.$$

I confess, to my shame, that I have not been able to make much progress towards solving these equations, or discovering their geometrical significance. In spite of that I still think that the method here outlined is the most promising for studying the problems indicated.

HARVARD UNIVERSITY,  
CAMBRIDGE, MASS.

# ON CIRCAVARIANT MATRICES AND CIRCA-EQUIVALENT NETWORKS

BY

RICHARD STEVENS BURINGTON

**1. Introduction.** In recent papers<sup>(1)</sup> the author has considered various questions concerning the equivalence of quadrics in  $m$ -affine  $n$ -space and related problems in the theory of (absolutely) equivalent  $m$ -terminal pair electrical networks.

The present paper is concerned with the development of certain theorems relating to the theory of congruent matrices which appear to be fundamental to the construction of a somewhat more general theory of (relative) equivalent electrical networks.

Consider the set of matrices  $B$  congruent to the matrix  $A$ ; i.e.,  $B = P'AP$ . In the first section of this paper a theory of circavariant matrices is initiated, general theorems being obtained relating to the restrictions which must be imposed on  $P$  in order that one or more of a certain set  $A_1, A_1^2, \dots$  of matrices derived from  $A$  each be circavariant. In later sections theorems on the congruence of matrices with  $P$  in a modified  $m$ -affine space are obtained, together with a set of normal forms.

In the last section, the theory of circavariant matrices is used to initiate a general theory of circa-equivalent networks, the usual theory of equivalent networks appearing as a special case of the general theory.

**2. Congruent and circavariant matrices.** Let  $A, B, C, \dots, P, Q, \bar{A}, \dots$  be matrices of order  $n$  whose elements belong to a field  $\mathfrak{F}$ . The matrix  $B$  is said to be *equivalent*<sup>(2)</sup> to the matrix  $A$  if there exist nonsingular matrices  $P$  and  $Q$  such that  $B = QAP$ . The matrix  $B$  is said to be *congruent* to  $A$  if there exists a nonsingular matrix  $P$  such that  $B = P'AP$ .

Let  $C_{r_1 \dots r_i}^{s_1 \dots s_i}$  denote the matrix obtained from  $C$  by deleting from  $C$  rows  $r_1, \dots, r_i$  and columns  $s_1, \dots, s_i$ . Denote  $C_{r_1 \dots r_i}^{s_1 \dots s_i}$  by  $C_{r_1 \dots r_i}$ .

Consider the set  $\mathfrak{A}$  of all matrices  $A$  of order  $n$  whose elements range

Presented to the Society, November 25, 1938, under the title *On the congruence of matrices and associated circavariant matrices*; received by the editors December 4, 1939.

<sup>(1)</sup> Burington, Richard S., *On the equivalence of quadrics in  $m$ -affine  $n$ -space and its relation to the equivalence of  $2m$ -pole networks*, these Transactions, vol. 38 (1935), pp. 163-176; hereafter called paper [1].

Burington, Richard S., *Matrices in electric circuit theory*, Journal of Mathematics and Physics, vol. 14 (1935), pp. 235-249; hereafter called paper [2].

Burington, Richard S., *R-matrices and equivalent networks I*, Journal of Mathematics and Physics, vol. 16 (1937), pp. 85-103; hereafter called paper [3].

<sup>(2)</sup> Throughout this paper it is understood, unless otherwise stated, that all the elements of all matrices used belong to a commutative field  $\mathfrak{F}$  whose characteristic is not two.

over  $\mathfrak{F}$ . With each  $A$  associate the set  $\mathfrak{B}$  of all matrices congruent to  $A$ . If  $B$  is any matrix of set  $\mathfrak{B}$ , there exists a nonsingular matrix  $P$  such that

$$(2.1) \quad B = P'AP.$$

If  $T$  ranges over the set  $\mathfrak{P}$  of all nonsingular matrices of order  $n$ , then the matrices  $\beta = T'AT$  are all congruent to  $A$ , and hence, each  $\beta$  belongs to  $\mathfrak{B}$ . The matrix  $P$  in (2.1) belongs to  $\mathfrak{P}$ . If (2.1) holds and if there exists a subset  $\mathfrak{P}_c$  of  $\mathfrak{P}$  such that for all matrices  $A$  of  $\mathfrak{A}$ , and for all matrices  $P$  of  $\mathfrak{P}_c$ ,

$$(2.2) \quad B_{r_1 \dots r_i}^{s_1 \dots s_i} = P_{r_1 \dots r_i}^{r'_1 \dots r'_i} A_{r'_1 \dots r'_i}^{s'_1 \dots s'_i} P_{s'_1 \dots s'_i}^{s_1 \dots s_i},$$

then  $A_{r_1 \dots r_i}^{s_1 \dots s_i}$  is called a *circavariant matrix* of  $A$  under the congruence (2.1). Let  $\mathfrak{B}_c$  denote the subset of  $\mathfrak{B}$  obtained by letting  $P$  range over  $\mathfrak{P}_c$ .

Thus, if (2.1) holds and if  $A_1$  is a circavariant matrix, then  $B_1$  can be obtained directly from the product  $B_1 = P'_1 A_1 P_1$ , or from  $B$  by deleting the first row and column.

The term *circavariant matrix* has been introduced here to avoid confusion with the term *invariant matrix* as used by L. Schur, Littlewood and other writers, as in D. E. Littlewood, *The construction of invariant matrices*, Proceedings of the London Mathematical Society, (2), vol. 43 (1937), pp. 226-240. In paper [1], a *circavariant matrix* was called an *invariant matrix*. In contrast to the definition used in [1], the present definition places greater emphasis on the requirement that (2.2) hold for all matrices of  $\mathfrak{A}$ . While in [1]  $P$  was restricted to (simply)  $m$ -affine types, here  $P$  is not so constrained.

3. Conditions that  $A_{r_1 \dots r_i}^{s_1 \dots s_i}$  be circavariant. In paper [1] a system of integer, matrix and algebraic invariants of the matrix  $A$  of the  $n$ -ary quadratic form  $F$  was exhibited, under the simply  $m$ -affine nonsingular group of linear transformations  $T$ , by means of which necessary and sufficient conditions for the simply  $m$ -affine congruence with respect to  $T$  of two matrices  $A$  and  $B$ , as well as the equivalence of the two corresponding forms  $F$  and  $G$ , were given.

Whereas in paper [1] we were concerned with the nature of the matrix  $A$  for given simply  $m$ -affine matrices  $P$ , in the present paper we are concerned as to the content of the subset  $\mathfrak{P}_c$ , that is, as to the conditions which must be imposed on  $P$  in order that  $A_{r_1 \dots r_i}^{s_1 \dots s_i}$  be a circavariant matrix of  $A$  for the class  $\mathfrak{A}$  under (2.1). We shall see that the solution to this question leads to a more general type of matrix  $P$  than that used in paper [1].

To begin with, we search for conditions on  $P$  in order that  $A_1$  be a circavariant matrix. In other words, with reference to congruence (2.1), under what conditions is the matrix  $M = P'_1 A_1 P_1$  identically equal to  $B_1$  in the elements of  $A$ ?

For convenience, we number the first row (and column) of  $A_1$  ( $P_1$ ,  $M$  and  $B_1$ ) as 2, the second as 3,  $\dots$ , the  $(n-1)$ -th as  $n$ . The rows (and columns) of  $A$  ( $P$  and  $B$ ) are numbered in the usual way, the first row as 1, the second row as 2,  $\dots$ , the  $n$ th row as  $n$ . Evidently,



$$(3.1) \quad M = \left( \sum_{r=2}^n \sum_{s=2}^n p_{rj} a_{rs} p_{sk} \right) = (m_{jk}),$$

where the element in the  $j$ th row and  $k$ th column of  $M$  is  $m_{jk}$ , with rows and columns numbered (as agreed above)  $j, k = 2, \dots, n$ . Also from (2.1)

$$(3.2) \quad B = \left( \sum_{r=1}^n \sum_{s=1}^n p_{rj} a_{rs} p_{sk} \right) = (b_{jk}),$$

where  $b_{jk}$  is the element in the  $j$ th row and  $k$ th column, ( $j, k = 1, 2, \dots, n$ ). In order that  $B_1$  be identical to  $M$  in the elements of  $A$ , we must have

$$(3.3) \quad b_{jk} = \sum_{r=1}^n \sum_{s=1}^n p_{rj} a_{rs} p_{sk} = \sum_{r=2}^n \sum_{s=2}^n p_{rj} a_{rs} p_{sk} = m_{jk},$$

in the elements of  $A$ , for  $j$  and  $k$  any fixed pair of integers selected from  $2, \dots, n$ . From (3.3) with  $j$  and  $k$  fixed, ( $j, k = 2, 3, \dots, n$ ), we find that the following identities in the elements of  $A$  must hold,

$$(3.4) \quad \begin{aligned} p_{1j} a_{1s} p_{sk} &= 0 & (s = 1, 2, \dots, n), \\ p_{rj} a_{r1} p_{1k} &= 0 & (r = 1, 2, \dots, n). \end{aligned}$$

The cases  $j=k$  with  $s=1$  give  $p_{1j} a_{11} p_{1j} = 0$  in  $a_{11}$ , ( $j=2, 3, \dots, n$ ), hence  $p_{1j}$  must vanish for  $j=2, 3, \dots, n$ . With  $p_{1j}=0$ , ( $j=2, \dots, n$ ), all the identities (3.4) are satisfied and  $A_1$  is a circavariant matrix. Since  $P$  is nonsingular,  $p_{11} \neq 0$ . Hence

**THEOREM 3.1.** *A necessary and sufficient condition that  $A_1$  be a circavariant matrix is that  $p_{1j}=0$ ,  $j=2, \dots, n$ .*

A matrix of the form

$$(3.5) \quad S = \begin{pmatrix} p_{11} & 0 & \cdots & 0 & \vdots & 0 & \cdots & 0 \\ 0 & p_{22} & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \cdots & p_{mm} & \vdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \vdots & \cdots & \cdots & \cdots \\ p_{m+1,1} & \cdots & p_{m+1,m} & \vdots & p_{m+1,m+1} & \cdots & p_{m+1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n,1} & \cdots & p_{n,m} & \vdots & p_{n,m+1} & \cdots & p_{nn} \end{pmatrix},$$

with elements in  $\mathfrak{F}$  is said to be  $m$ -affine<sup>(\*)</sup>. If  $p_{11}=p_{22}=\cdots=p_{mm}=1$ ,  $S$  is said to be *simply  $m$ -affine*.

(\*) In paper [1], the term  $m$ -affine means *simply  $m$ -affine*.

The matrix  $B$  is said to be  $m$ -affine congruent to  $A$  if there exists an  $m$ -affine nonsingular matrix  $S$  with elements in  $\mathfrak{F}$  such that  $B = S'AS$ .

Theorem 3.1 states that a necessary and sufficient condition that  $A_1$  be circavariant is that  $P$  be 1-affine. If  $B$  is 1-affine congruent to  $A$ , then  $A_1$  is circavariant and  $B_1$  is congruent to  $A_1$ .

More generally, suppose that we require that  $A_u$  be circavariant. Then for  $j$  and  $k$  any fixed pair of integers selected from  $1, 2, \dots, u-1, u+1, \dots, n$  the following identity in the elements of  $A$  must hold:

$$(3.6) \quad \sum_{r=1}^n \sum_{s=1}^n p_{rj} a_{rs} p_{sk} \equiv \sum_{r=1, r \neq u}^n \sum_{s=1, s \neq u}^n p_{rj} a_{rs} p_{sk}.$$

This means that the following identities in the elements of  $A$  must hold:

$$(3.7) \quad \begin{aligned} p_{uj} a_{us} p_{sk} &\equiv 0 & (s = 1, 2, \dots, n), \\ p_{rj} a_{ru} p_{uk} &\equiv 0 & (r = 1, 2, \dots, n). \end{aligned}$$

The cases  $j = k \neq u$  with  $s = u$  give  $p_{uj} a_{uu} p_{uj} = 0$ , so that each  $p_{uj} = 0, j \neq u$ . We conclude

**THEOREM 3.2.** *A necessary and sufficient condition that  $A_u$  be circavariant is that  $p_{uj} = 0$  for  $j = 1, \dots, n; j \neq u$ .*

We note that  $d(P) = p_{uu} \cdot d(P_u)$ . Since  $P$  is nonsingular  $p_{uu} \neq 0, d(P_u) \neq 0$ . From (2.2), we conclude that  $B_u$  is congruent to  $A_u$ . Hence if  $A_u$  is circavariant,  $B_u$  is congruent to  $A_u$ .

Suppose we require  $A_v^u, u \neq v$ , to be circavariant. Then for  $j$  and  $k$  any fixed pair of integers selected from  $j = 1, \dots, u-1, u+1, \dots, n$  and  $k = 1, \dots, v-1, v+1, \dots, n$ , we must have

$$\sum_{r=1}^n \sum_{s=1}^n p_{rj} a_{rs} p_{sk} \equiv \sum_{r=1, r \neq u}^n \sum_{s=1, s \neq v}^n p_{rj} a_{rs} p_{sk}$$

identically in the elements of  $A$ ; i.e.,

$$(3.8) \quad \begin{aligned} p_{uj} a_{us} p_{sk} &\equiv 0 & (s = 1, \dots, n), \\ p_{rj} a_{rv} p_{vk} &\equiv 0 & (r = 1, \dots, n). \end{aligned}$$

The cases  $j = k = w, w \neq u, w \neq v$ , with  $s = u$  and  $r = v$  give  $p_{uw} a_{uu} p_{uw} = 0$  and  $p_{vw} a_{vv} p_{vw} = 0$ , so that  $p_{uw} = p_{vw} = 0$ . The case  $j = v, k = u$  gives

$$(3.9) \quad \begin{aligned} p_{uv} a_{us} p_{su} &\equiv 0 & (s = 1, \dots, n), \\ p_{rv} a_{rv} p_{vu} &\equiv 0 & (r = 1, \dots, n). \end{aligned}$$

Since  $P$  is nonsingular at least one  $p_{su} \neq 0$ . Hence  $p_{uv} = 0$ . Likewise, at least one  $p_{rv} \neq 0$ , so that  $p_{vu} = 0$ . The case  $A_v^u$  leads to the same result. We have

**THEOREM 3.3.** *A necessary and sufficient condition that  $A_u^*$ ,  $u \neq v$ , be a circavariant matrix is that*

$$\begin{aligned} p_{u\alpha} &= 0 & (\alpha = 1, \dots, n, \alpha \neq u); \\ p_{v\beta} &= 0 & (\beta = 1, \dots, n, \beta \neq v). \end{aligned}$$

This theorem shows that a necessary and sufficient condition that  $A_1^2$  be circavariant is that  $P$  be 2-affine.

Evidently  $d(P) = p_{uu}p_{vv} \cdot d(P_{uv})$ . Since  $P$  is nonsingular,  $p_{uu}p_{vv} \neq 0$ ,  $d(P_{uv}) \neq 0$ . Hence from (2.2),  $B_u^*$  is equivalent to  $A_u^*$ . Thus,  $B_u^*$  is equivalent to  $A_u^*$  when  $A_u^*$  is circavariant.

If  $A_{uv}$ ,  $u \neq v$ , is a circavariant matrix, then for  $j$  and  $k$  any fixed numbers selected from  $w = 1, \dots, n$ ,  $w \neq u$ ,  $w \neq v$ , we must have

$$\sum_{r=1}^n \sum_{s=1}^n p_{rj} a_{rs} p_{sk} = \sum_{r=1, r \neq u, v}^n \sum_{s=1, s \neq u, v}^n p_{rj} a_{rs} p_{sk}$$

identically in the elements of  $A$ . This means that

$$(3.10) \quad \begin{aligned} p_{uj} a_{us} p_{sk} &\equiv 0, & p_{vj} a_{vs} p_{sk} &\equiv 0 & (s = 1, 2, \dots, n), \\ p_{rj} a_{ru} p_{uk} &\equiv 0, & p_{rj} a_{rv} p_{vk} &\equiv 0 & (r = 1, 2, \dots, n) \end{aligned}$$

identically in the elements of  $A$ . The cases  $j = k = w = 1, \dots, n$ ,  $w \neq u$ ,  $w \neq v$ , with  $s = u, v$  give  $p_{uw} a_{uu} p_{uw} = 0$ ,  $p_{vw} a_{vv} p_{vw} = 0$ . We conclude that  $p_{uw} = p_{vw} = 0$ . Since  $P$  is nonsingular,  $p_{uu}p_{vv} - p_{uv}p_{vu} \neq 0$ , and  $d(P_{uv}) \neq 0$ .

**THEOREM 3.4.** *A necessary and sufficient condition that  $A_{uv}$  be a circavariant matrix is that  $p_{uw} = p_{vw} = 0$ , for  $w = 1, \dots, n$ ,  $w \neq u$ ,  $w \neq v$ .*

This theorem shows that a sufficient (but not necessary) condition that  $A_{12}$  be circavariant is that  $P$  be 2-affine.

Let  $J$  be the set  $1, \dots, n$  and let  $u_1, u_2, \dots, u_g$  be any subset  $U$  of  $J$ , all the elements of  $U$  being distinct. Denote by  $W = J - U$  the set  $J$  with the elements of  $U$  removed. If  $r$  is not in  $W$ , we write  $r \notin W$ .

If  $A_{u_1 \dots u_g}$  is a circavariant matrix, then for  $j$  and  $k$  any fixed numbers selected from  $W$ , we must have

$$\sum_{r=1}^n \sum_{s=1}^n p_{rj} a_{rs} p_{sk} = \sum_{r=1, r \notin W}^n \sum_{s=1, s \notin W}^n p_{rj} a_{rs} p_{sk},$$

identically in the elements of  $A$ . This means that

$$(3.11) \quad \begin{aligned} p_{u_1 j} a_{u_1 s} p_{sk} &\equiv 0, \dots, p_{u_g j} a_{u_g s} p_{sk} \equiv 0 & (s = 1, 2, \dots, n), \\ p_{rj} a_{ru_1} p_{u_1 k} &\equiv 0, \dots, p_{rj} a_{ru_g} p_{u_g k} \equiv 0 & (r = 1, 2, \dots, n), \end{aligned}$$

identically in the elements of  $A$ . The cases where  $j = k$  and  $j$  ranges over  $W$  with  $s = u_1, u_2, \dots, u_g$  give

$$p_{u_1 j} a_{u_1 u_1} p_{u_1 j} \equiv 0, p_{u_2 j} a_{u_2 u_2} p_{u_2 j} \equiv 0, \dots, p_{u_g j} a_{u_g u_g} p_{u_g j} \equiv 0,$$

from which we conclude that  $p_{u_1j} = p_{u_2j} = \dots = p_{u_gj} = 0$ , for  $j$  ranging over  $W$ .

**THEOREM 3.5.** *A necessary and sufficient condition that  $A_{u_1 \dots u_g}$  be circavariant is that  $p_{u_1j} = p_{u_2j} = \dots = p_{u_gj} = 0$  for  $j$  ranging over  $W$ .*

From  $P$  delete all the rows and columns whose numbers belong to the set  $W$ , leaving the matrix  $P_W$ . It is easy to see that  $d(P) = d(P_W) \cdot d(P_{u_1 \dots u_g})$ . Since  $d(P) \neq 0$ ,  $P_W$  and  $P_{u_1 \dots u_g}$  are both nonsingular, so that  $B_{u_1 \dots u_g}$  is congruent to  $A_{u_1 \dots u_g}$  when the latter is circavariant.

Let  $(u_1, u_2, \dots, u_i)$  and  $(v_1, v_2, \dots, v_i)$  be two subsets  $U$  and  $V$ , respectively, of the set  $I$  of integers  $1, 2, \dots, n$ . Suppose that all the elements of  $U$  and  $V$  are distinct. Let  $L = U - V$  be the set  $I$  with the elements of  $U$  and  $V$  removed. Then

**THEOREM 3.6.** *A necessary and sufficient condition that  $A_{u_1 \dots u_i}^{v_1 \dots v_i}$  (or  $A_{v_1 \dots v_i}^{u_1 \dots u_i}$ ) be circavariant is that  $P$  be such that for each  $u$  in  $U$ , each  $v$  in  $V$ , and for each  $\lambda$  in  $L$ ,  $p_{u\lambda} = p_{uv} = p_{vu} = 0$ .*

The proof of this theorem may be obtained from the proof of Theorem 3.3 as follows: equations (3.8) must hold with  $u$  ranging over  $U$  and  $v$  ranging over  $V$ . The cases  $j = k = \lambda$  with  $s = u$ ,  $r = v$ ,  $u$  ranging over  $U$  and  $v$  over  $V$ , lead to the conditions  $p_{u\lambda} = 0$ . The cases  $j = v$ ,  $k = u$ , with  $u$  over  $U$  and  $v$  over  $V$  yield  $p_{uv} = 0$  and  $p_{vu} = 0$ .

It is easy to see that  $B_{u_1 \dots u_i}^{v_1 \dots v_i}$  is equivalent to  $A_{u_1 \dots u_i}^{v_1 \dots v_i}$  in case the latter is circavariant. For example, if  $A_{14}^{23}$  is circavariant,  $P$  is such that

$$d(P) = d(E) \cdot d(F) \cdot d(P_{12}^{34})$$

where

$$E = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} p_{33} & p_{34} \\ p_{43} & p_{44} \end{pmatrix}.$$

Since  $P$  is nonsingular so are  $E$ ,  $F$ , and  $P_{12}^{34}$ . From (2.2),  $B_{12}^{34}$  is equivalent to  $A_{12}^{34}$ .

In a similar manner further theorems concerning the circavariance of  $A_{u_1 \dots u_i}^{v_1 \dots v_i}$  for the case when the sets  $U$  and  $V$  overlap can be stated together with theorems concerning the equivalence of  $B_{u_1 \dots u_i}^{v_1 \dots v_i}$  and  $A_{u_1 \dots u_i}^{v_1 \dots v_i}$ .

**4. Invariants.** From (2.1),  $d(B) = [d(A)] [d(P)]^2$ . Hence, the determinant of  $A$  is a relative invariant of the set  $\mathfrak{B}$  under (2.1) with  $P$  ranging over the set  $\mathfrak{P}$ . Since each  $P$  in  $\mathfrak{P}$  is nonsingular, the rank of  $B$  is equal to the rank of  $A$ , so that the rank of  $A$  is an integer invariant for the set  $\mathfrak{B}$ . If the field  $\mathfrak{F}$  is ordered, the signatures (when defined) of  $B$  and  $A$  are equal, so that for ordered fields, the signature of  $A$  is an integer invariant for the set  $\mathfrak{B}$ .

Suppose in (2.2),  $P'_{1 \dots r}$  and  $P_{s_1 \dots s_r}$  are nonsingular and that  $A_{r_1 \dots r_i}^{s_1 \dots s_i}$  is a circavariant matrix. Consider any function  $G(a_{ij}) = G$  of the elements of  $A_{r_1 \dots r_i}^{s_1 \dots s_i}$  which is so related to the same function  $G(b_{ij}) = \bar{G}$  of the elements of

$B_{r_1 \dots r_t}^{s_1 \dots s_t}$  that, in the elements  $a_{ij}$ ,

$$(4.1) \quad \bar{G} = \alpha G \beta \quad (\alpha \beta \neq 0),$$

where  $\alpha = \alpha(P'_{r_1 \dots r_t})$  is a function of the elements of  $P'_{r_1 \dots r_t}$  only, and where  $\beta = \beta(P_{s_1 \dots s_t})$  is a function of the elements of  $P_{s_1 \dots s_t}$  only. Then  $G$  is said to be a *circavariant of the set  $\mathfrak{B}$  with respect to  $A_{r_1 \dots r_t}^{s_1 \dots s_t}$* .

If  $G, G_1, G_2, \dots, G_{r_1 \dots r_t}^{s_1 \dots s_t}$  are circavariants of the set  $\mathfrak{B}$  with respect to the circavariant matrices  $A, A_1, A_2, \dots, A_{r_1 \dots r_t}^{s_1 \dots s_t}$ , respectively, then any function  $H$  of the form

$$(4.2) \quad H(G) = [G]^{\rho_1} [G_1]^{\rho_2} [G_2]^{\rho_3} \dots [G_{r_1 \dots r_t}^{s_1 \dots s_t}]^{\rho_n},$$

where the  $\rho_i$ 's are real numbers is said to be a *composite circavariant* of the set  $\mathfrak{B}$ . Let  $H(\bar{G}) = \bar{H}$  denote  $H$  with  $G, G_1, \dots, G_{r_1 \dots r_t}^{s_1 \dots s_t}$  replaced by  $\bar{G}, \bar{G}_1, \bar{G}_2, \dots, \bar{G}_{r_1 \dots r_t}^{s_1 \dots s_t}$ , respectively. Then  $\bar{H}$  is of the form

$$(4.3) \quad \bar{H} = \gamma H \delta \quad (\gamma \delta \neq 0),$$

where

$$\gamma = [\alpha_1(P')] [\alpha_2(P'_1)] \dots [\alpha_n(P'_{r_1 \dots r_t})], \quad \delta = [\beta_1(P)] [\beta_2(P_1)] \dots [\beta_n(P_{s_1 \dots s_t})].$$

If  $\gamma \delta = 1$ , then  $H$  is said to be an *absolute circavariant* of  $\mathfrak{B}$ . If  $\gamma = \delta$ , then  $H$  is said to be a *relative circavariant* of  $\mathfrak{B}$ .

Consider the set  $\mathfrak{B}_{r_1 \dots r_t}^{s_1 \dots s_t}$  of all matrices  $B_{r_1 \dots r_t}^{s_1 \dots s_t}$  generated from the circavariant matrix  $A_{r_1 \dots r_t}^{s_1 \dots s_t}$  by letting  $P$  range over  $\mathfrak{P}_s$ , with  $P'_{r_1 \dots r_t}$  and  $P_{s_1 \dots s_t}$  nonsingular. Then the rank  $\rho_{r_1 \dots r_t}^{s_1 \dots s_t}$  of each matrix in  $\mathfrak{B}_{r_1 \dots r_t}^{s_1 \dots s_t}$  is equal to the rank of  $A_{r_1 \dots r_t}^{s_1 \dots s_t}$ .

We suppose that  $(r_1, \dots, r_t) = (s_1, \dots, s_t)$ . Then (2.2) is an ordinary congruence. Suppose the field  $\mathfrak{F}$  is ordered and that a  $P$  exists in  $\mathfrak{P}_s$  for which  $B_{r_1 \dots r_t}$  is a diagonal matrix, so that  $A_{r_1 \dots r_t}$  has a signature  $\sigma_{r_1 \dots r_t}$ . From (2.2) it follows that the signature of each matrix in the set  $\mathfrak{B}_{r_1 \dots r_t}$  is equal to  $\sigma_{r_1 \dots r_t}$ . Thus,

**THEOREM 4.1.** *The rank of  $A_{r_1 \dots r_t}^{s_1 \dots s_t}$  is an integer invariant for the set  $\mathfrak{B}_{r_1 \dots r_t}^{s_1 \dots s_t}$ . If  $\mathfrak{F}$  is ordered, the signature (when defined) of  $A_{r_1 \dots r_t}$  is an integer invariant for the set  $\mathfrak{B}_{r_1 \dots r_t}$ .*

If  $A_{r_1 \dots r_t}^{s_1 \dots s_t}$  is a circavariant matrix, then from (2.2)

$$(4.4) \quad d(B_{r_1 \dots r_t}^{s_1 \dots s_t}) = d(P'_{r_1 \dots r_t}) \cdot d(A_{r_1 \dots r_t}^{s_1 \dots s_t}) \cdot d(P_{s_1 \dots s_t}).$$

From (4.4) it is clear that  $d(A_{r_1 \dots r_t}^{s_1 \dots s_t})$  is a circavariant for the set  $\mathfrak{B}$ . If  $(r_1, \dots, r_t) = (s_1, \dots, s_t)$ ,

$$(4.5) \quad d(B_{r_1 \dots r_t}^{r_1 \dots r_t}) = [d(P_{r_1 \dots r_t})]^2 [d(A_{r_1 \dots r_t}^{r_1 \dots r_t})],$$

so that



**THEOREM 4.2.** *The determinants of the matrices of the set  $\mathfrak{B}_{r_1 \dots r_t}^{s_1 \dots s_t}$  are circavariants for the set  $\mathfrak{B}$  with respect to  $A_{r_1 \dots r_t}^{s_1 \dots s_t}$ . If  $(r_1, \dots, r_t) \equiv (s_1, \dots, s_t)$ , these determinants are relative circavariants of  $\mathfrak{B}$ .*

It may be remarked that in case  $(r_1, \dots, r_t) \equiv (s_1, \dots, s_t)$  these determinants are actually ordinary relative invariants of  $\mathfrak{B}_{r_1 \dots r_t}^{s_1 \dots s_t}$  under an ordinary congruence of transformation matrix  $P_{r_1 \dots r_t}$ .

Evidently the ratio of any two circavariants is a composite circavariant for the set  $\mathfrak{B}$ .

**5. Normal forms for  $A$  under  $P$   $m$ -affine.** In the theory of electrical networks the cases when  $A_1, A_1^2, A_2, \dots$  are to be circavariant frequently occur, leading to the requirement that  $P$  be  $m$ -affine. We shall accordingly consider the normal forms of  $A$  under  $P$   $m$ -affine.

In paper [1] the reduction of  $A$  to normal forms was indicated, the case where  $m=2$  being considered in detail. In particular, the results obtained indicate that, when  $P$  is simply  $m$ -affine, every symmetric matrix  $A$  with elements in a field  $\mathfrak{F}$  (not of characteristic two) with circavariant matrix  $A_1, \dots, A_{m-1}$  is  $m$ -affine congruent in  $\mathfrak{F}$  to a matrix  $B$  in which the matrix  $B_1, \dots, B_{m-1}$  is of the form

$$(5.1) \quad \begin{pmatrix} b_{mm} & 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & \cdot & b_{m+1} & 0 & \dots & 0 \\ 0 & \cdot & 0 & b_{m+2} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & \cdot & 0 & \cdot & \dots & b_{n-1} & 0 \\ 0 & \cdot & 0 & 0 & \dots & 0 & b_n \end{pmatrix}, \quad \text{if } \nu = \rho_1 \dots m-1 - \rho_1 \dots m \neq 2,$$

and with a parabolic matrix

$$(5.2) \quad \begin{pmatrix} 0 & \cdot & 0 & 0 & \dots & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & \cdot & b_{m+1} & 0 & \dots & 0 & 0 \\ 0 & \cdot & 0 & b_{m+2} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & 0 & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & \cdot & 0 & 0 & \dots & b_{n-1} & 0 \\ 1 & \cdot & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad \text{if } \nu = 2,$$

the number of nonzero  $b_i$ 's in  $B_1, \dots, B_m$  being equal to the rank  $\rho_1 \dots m$  of  $A_1, \dots, A_m$ .

The parameter  $b_{mm}$  is an absolute invariant of  $A$  when  $P$  is simply  $m$ -affine.

We shall let  $\sigma_{1\dots m} = \sigma_{1\dots m}^m$  denote the signature of  $A_{1\dots m}$ . If the field  $\mathfrak{F}$  is real, each positive  $b_i$  in  $B_{1\dots m}$  can be reduced (by means of a simply  $m$ -affine  $P$ ) to 1, and each negative  $b_i$  to  $-1$ . The number of positive  $b_i$ 's in  $B_{1\dots m}$  is  $(\rho_{1\dots m} + \sigma_{1\dots m})/2$  and the number of negative  $b_i$ 's is  $(\rho_{1\dots m} - \sigma_{1\dots m})/2$ , the remaining  $b_i$ 's each being zero. If  $\mathfrak{F}$  is algebraically closed, each nonzero  $b_i$  in  $B_{1\dots m}$  can be reduced to 1. No further reduction of  $B_{1\dots m-1}$  is possible when  $P$  is simply  $m$ -affine. For (5.1) and (5.2), we shall denote the reduced form of  $B_{1\dots m}$  thus obtained by  $\delta = [\delta_{m+1}, \dots, \delta_{r-1}, 0, \dots, 0]$ , a diagonal matrix.

In case  $\mathfrak{F}$  is ordered, we shall agree to *regularly arrange*<sup>(4)</sup> the matrix  $\delta$ , this always being possible when  $P$  is  $m$ -affine.

Suppose in (5.1),  $b_{mm} \neq 0$ . Let  $P$  be  $m$ -affine with  $p_{rs} = \delta_{rs}$ , (where  $\delta_{rs} = 0$  if  $r \neq s$ ,  $\delta_{rs} = 1$  if  $r = s$ ), except for  $p_{mm}$ . If  $\mathfrak{F}$  is algebraically closed, select a  $p_{mm}$  so that  $p_{mm}^2 = 1/b_{mm}$ . If  $\mathfrak{F}$  is real, let  $p_{mm} = 1/(b_{mm})^{1/2}$  if  $b_{mm} > 0$  and  $p_{mm} = 1/(-b_{mm})^{1/2}$  if  $b_{mm} < 0$ . Then, in case of (5.1) with  $b_{mm} \neq 0$ , the matrix  $B$  is  $m$ -affine congruent to a matrix  $C = P'BP$  in which the matrix  $C_{1\dots, m-1}$  is of the form

$$(5.3) \quad \begin{pmatrix} b_m & 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & \delta_{m+1} & 0 & \dots & 0 \\ 0 & 0 & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \delta_{r-1} & \cdot \\ \cdot & \cdot & \cdot & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \dots & 0 \end{pmatrix}, \quad \text{if } r \neq 2,$$

where  $b_m$  is 1 if  $\mathfrak{F}$  is algebraically closed; and  $b_m = 1$  or  $-1$  if  $\mathfrak{F}$  is real. In the latter case,  $b_{mm} = 1$  when  $\sigma_{1\dots m-1} = 1 + \sigma_{1\dots m}$ ,  $b_{mm} = -1$  when  $\sigma_{1\dots m-1} = -1 + \sigma_{1\dots m}$ , and  $b_{mm} = 0$  when  $\sigma_{1\dots m-1} = \sigma_{1\dots m}$ .

As in paper [1], it is now easy to formulate various theorems. For example, Theorem 3.3 of [1] for  $P$   $m$ -affine holds without the requirement on the parameters  $b_{mm}$  and  $b'_{mm}$ .

**THEOREM 5.1.** *Let  $P$  be  $m$ -affine with elements in an algebraically closed field  $\mathfrak{F}$ , and let  $A^{(1)}$  and  $A^{(2)}$  be two symmetric matrices of order  $n$  in  $\mathfrak{F}$  with associated circavariant matrices  $A_{1\dots, m-1}^{(1)}$  and  $A_{1\dots, m-1}^{(2)}$ . A necessary and sufficient condition that  $A_{1\dots, m-1}^{(1)}$  and  $A_{1\dots, m-1}^{(2)}$  be congruent is that they have the*

<sup>(4)</sup> C. C. MacDuffee, *Theory of Matrices*, Berlin, 1933, pp. 57-58.

same ranks  $\rho_{1,\dots,m-1}^{(1)}$ ,  $\rho_{1,\dots,m}^{(1)}$ , and  $\rho_{1,\dots,m-1}^{(2)}$ ,  $\rho_{1,\dots,m}^{(2)}$ , respectively. If the field  $\mathfrak{F}$  is real the additional requirement of the equality of the signatures  $\sigma_{1,\dots,m-1}^{(1)}$ ,  $\sigma_{1,\dots,m}^{(1)}$  and  $\sigma_{1,\dots,m-1}^{(2)}$ ,  $\sigma_{1,\dots,m}^{(2)}$ , respectively, must be met.

Case  $m=2$ . If  $m=2$ , it was shown in [1] that the symmetric matrix  $A$  is simply 2-affine congruent to one of the various normal forms  $f_1, f_2, \dots, f_5$  given below and as indicated in Table I ( $\delta$  regularly arranged):

$$\begin{aligned}
 f_1 &= \begin{pmatrix} 0 & 0 & \cdot & 0 & \cdots & 0 & 1 \\ 0 & b_{22} & \cdot & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdot & & & & \delta \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 0 & 0 & \cdot & & & & \cdot \\ 1 & 0 & \cdot & & & & \cdot \end{pmatrix}, & f_2 &= \begin{pmatrix} b_{11} & b_{12} & \cdot & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdot & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdot & & & \delta \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ 0 & 0 & \cdot & & & \cdot \end{pmatrix}, \\
 f_3 &= \begin{pmatrix} 0 & 0 & \cdot & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdot & 0 & \cdots & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & & & & & \delta \\ \cdot & \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & \cdot & & & & & \cdot \\ 0 & 0 & \cdot & & & & & \cdot \\ 1 & 0 & \cdot & & & & & \cdot \\ 0 & 1 & \cdot & & & & & \cdot \end{pmatrix}, & f_4 &= \begin{pmatrix} b_{11} & 0 & \cdot & 0 & \cdots & 0 & b_{1n} \\ 0 & 0 & \cdot & 0 & \cdots & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdot & & & & \delta \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 0 & 0 & \cdot & & & & \cdot \\ b_{1n} & 1 & \cdot & & & & \cdot \end{pmatrix}, \\
 f_5 &= \begin{pmatrix} b_{11} & 0 & \cdot & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdot & 0 & \cdots & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdot & & & & \delta \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 0 & 0 & \cdot & & & & \cdot \\ 0 & 1 & \cdot & & & & \cdot \end{pmatrix}, & \delta &= \begin{pmatrix} \delta_3 & 0 & \cdot & \cdots & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \delta_{r-1} & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}.
 \end{aligned}$$

If  $P$  is 2-affine and nonsingular, it is possible through the proper selection of  $p_{11}$  and  $p_{22}$  to reduce certain of the matrices  $f_1, f_2, \dots, f_5$  yielding new matrices  $g_1, g_2, \dots, g_5$  having the same general form as  $f_1, \dots, f_5$ , respectively, each  $g_i$  being 2-affine congruent to  $f_i$  ( $i=1, \dots, 5$ ). The forms  $g_1, \dots, g_5$  are indicated in Table I. Thus, in the case of  $f_1$ , with  $\rho_1=r-2$ ,  $g_1$  is  $f_1$  with  $b_{22}$  replaced by 1 if  $\mathfrak{F}$  is algebraically closed, and with  $b_{22}$  replaced by 1

or  $-1$  if  $\mathfrak{F}$  is real. No further reduction of  $f_1$  is possible by a 2-affine  $P$  which preserves the form  $g_1$ . The numbers  $\delta_{11}$ ,  $\delta_{22}$  in Table I denote 1 if  $\mathfrak{F}$  is algebraically closed, and denote 1 or  $-1$  if  $\mathfrak{F}$  is real. In Case 3, the parameter  $b_{22}$  is an absolute invariant. It should be noted that the number of such parameters appearing in the  $g_i$ 's is just one, whereas in the simply 2-affine case there were several such parameters,  $b_{ii}$ , in the forms  $f_1, \dots, f_s$ . (See Table I, p. 171, paper [1], which may be constructed from Table I, as here given, by deleting the  $\delta$ 's and by replacing each 1 by the symbol  $\neq 0$ , and each  $f_i$  by  $g_i$ .)

TABLE I  
Classification of matrix  $A$  for the case  $m=2$

Case	$\rho_{12}=r-3, r=3, 4, \dots, n+1$					$P$ 2-affine					Form
	$\rho_1 - \rho_{12}$	$\rho_2 - \rho_{12}$	$\rho$	$\rho_1$	$\rho_2$	$b_1^2$	$b_{11}$	$b_{12}$	$b_{22}$	$b_{1n}$	
1	$\neq 2$	$= 2$		$r-2$					$\delta_{22} \neq 0$		$g_1$
2	$\neq 2$	$= 2$		$r-3$					0		$g_1$
3	$\neq 2$	$\neq 2$		$r-2$	$r-2$	$r-2$	$\delta_{11} \neq 0$	1	$b_{22} \neq 0$		$g_2$
4	$\neq 2$	$\neq 2$		$r-3$	$r-2$	$r-2$	$\delta_{11} \neq 0$	1	0		$g_2$
5	$\neq 2$	$\neq 2$		$r-3$	$r-2$	$r-3$	$\delta_{11} \neq 0$	0	0		$g_2$
6	$\neq 2$	$\neq 2$		$r-2$	$r-2$	$r-3$	$\delta_{11} \neq 0$	0	$\delta_{22} \neq 0$		$g_2$
7	$\neq 2$	$\neq 2$		$r-2$	$r-3$	$r-2$	0	1	$\delta_{22} \neq 0$		$g_2$
8	$\neq 2$	$\neq 2$		$r-3$	$r-3$	$r-2$	0	1	0		$g_2$
9	$\neq 2$	$\neq 2$		$r-3$	$r-3$	$r-3$	0	0	0		$g_2$
10	$\neq 2$	$\neq 2$		$r-2$	$r-3$	$r-3$	0	0	$\delta_{22} \neq 0$		$g_2$
11	$= 2$	$= 2$	$r+1$								$g_3$
12	$= 2$	$= 2$	$r$				$\delta_{11} \neq 0$			1	$g_4$
13	$= 2$	$= 2$	$r-1$				0			1	$g_4$
14	$= 2$	$\neq 2$	$r$		$r-2$		$\delta_{11} \neq 0$				$g_5$
15	$= 2$	$\neq 2$	$r-1$		$r-3$		0				$g_5$

If  $\mathfrak{F}$  is real, each form in Table I can be subdivided according to the signatures of  $A_{12}$ ,  $A_1$ ,  $A_2$ .

The following theorems are now evident:

**THEOREM 5.2.** *A symmetric matrix  $A$  with elements in a field  $\mathfrak{F}$  is 2-affine congruent in  $\mathfrak{F}$  to one of the forms  $g_1, \dots, g_6$ , according to the ranks (and signatures if  $\mathfrak{F}$  is real) of the circavariant matrices  $A$ ,  $A_1$ ,  $A_2$ ,  $A_1^2$ ,  $A_{12}$  as indicated in Table I.  $A$  is simply 2-affine congruent to one of the forms  $f_1, \dots, f_5$  as indicated in Table I.*

**THEOREM 5.3.** *A necessary and sufficient condition for the 2-affine congruence of two matrices  $A$  and  $C$  whose elements belong to the real field is that the circavariant matrices  $A$ ,  $A_1$ ,  $A_2$ ,  $A_1^2$ ,  $A_{12}$  and  $C$ ,  $C_1$ ,  $C_2$ ,  $C_1^2$ ,  $C_{12}$  have the same ranks and signatures, respectively, and that in Case 3 with  $P$  simply affine (Table I, paper [1]) the parameters  $b_{22}$  and  $b_{22}$ , for  $A$  and  $C$  respectively, be identically equal. If  $\mathfrak{F}$  is algebraically closed, the above holds without the signatures.*

*Case  $m = m$ .* The reduction of  $A$  to normal forms for  $P$   $m$ -affine may be done in a manner similar to that used above for the case when  $m = 2$ .

**6. Applications to the theory of forms and geometry.** It is a simple matter to translate the results of this paper into the language of the theory of bilinear forms under cogredient  $m$ -affine transformations.

A geometric study of the locus  $F = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = 0$  in a geometry built upon the group of transformations  $x_r = \sum_{s=1}^n p_{rs} y_s$ ,  $r = 1, \dots, n$ , with  $(p_{rs})$   $m$ -affine can also be made.

**7. Relation to linear networks** [2], [3]. We consider an  $m$ -terminal pair bilateral  $n$ -mesh linear electrical network  $\mathfrak{N}$  containing (lumped) resistances, inductances, and capacitances. Let  $E_1, \dots, E_m$  be the (complex) e.m.f.'s impressed on terminal pairs  $1, \dots, m$ , respectively;  $Q_s$  and  $I_s$ , the (complex) charge and (complex) current, respectively, in mesh  $s$ ;  $R_{st}$ ,  $L_{st}$ ,  $D_{st}$  (real numbers), the lumped circuit parameters (resistance, inductance, and elastance, respectively) for mesh  $s$  if  $s = t$ , common to meshes  $s$  and  $t$  if  $s \neq t$ . The meshes are so chosen that mesh  $s$ , ( $s = 1, \dots, m$ ), is the only one which passes through the terminal pair  $s$ .

Suppose the Kirchhof equations in complex form for the network are

$$(7.1) \quad B\{Q\} = \{E\},$$

where  $B = (b_{rs})$ ,  $b_{rs} = b_{sr} = L_{rs}\lambda^2 + R_{rs}\lambda + D_{rs}$ ,  $\lambda = j\omega$ ,  $\omega$  being the (real) frequency.

If  $B$  is nonsingular,

$$(7.2) \quad \{Q\} = B^{-1}\{E\}.$$

Let  $B^{-1} \equiv C = (c_{uv})$ . The element  $c_{uv}$  is called the *generalized (complex) network admittance*; being a transfer admittance between meshes  $u$  and  $v$  if  $u \neq v$ , and a driving-point admittance for mesh  $u$  if  $u = v$ .

Let  $Y = (Y_{st})$  be  $C$  with all but the first  $m$  rows and first  $m$  columns deleted. Then

$$(7.3) \quad \{Q\}_m = Y\{E\}_m,$$

where  $\{Q\}_m$  and  $\{E\}_m$  are the first  $m$  rows of  $\{Q\}$  and  $\{E\}$ , respectively. The matrix  $Y$  is called a *characteristic (admittance) coefficient matrix* for  $\mathfrak{N}$  [2].

Let  $\mathfrak{N}_1$  and  $\mathfrak{N}_2$  be two  $m$ -terminal pair networks of characteristic matrices  $Y^{(1)}$  and  $Y^{(2)}$ , respectively.  $\mathfrak{N}_2$  is said to be *circa-equivalent* to  $\mathfrak{N}_1$  if there exists a real nonsingular diagonal matrix  $D$  such that for all values of  $\lambda$ ,  $Y^{(1)} = D'Y^{(2)}D$ .

It should be noted here that this definition is much more general than the one heretofore used in the theory of equivalent electrical networks. The usual definition, the one given in papers [1], [2], and [3], is a very simple case of the one given in the present paper, being merely the type of circa-equivalence for which  $D$  is the identity matrix.



If  $\mathfrak{N}_1$  and  $\mathfrak{N}_2$  are circa-equivalent, then the admittances  $Y_{rs}^{(2)}$ , ( $r, s = 1, \dots, m$ ), are relative circavariants.

Let the diagonal element in the  $r$ th row of  $D$  be  $d_r$ , and let  $\Sigma = (\sigma_{rs})$  where  $\sigma_{rs} = d_r d_s$ . If all of the elements in row  $k$  of  $\Sigma$  are equal,  $\mathfrak{N}_2$  is said to be relatively equivalent to  $\mathfrak{N}_1$  with respect to terminal pair  $k$ . If each element in the  $k$ th row of  $\Sigma$  is equal to one,  $\mathfrak{N}_2$  is said to be absolutely equivalent to  $\mathfrak{N}_1$  with respect to terminal pair  $k$ . If all the elements of  $\Sigma$  are equal to a number  $\sigma$ ,  $\mathfrak{N}_2$  is relatively equivalent to  $\mathfrak{N}_1$ ; and in case  $\sigma = 1$ ,  $\mathfrak{N}_2$  is absolutely equivalent to  $\mathfrak{N}_1$ .

Consider the set  $\mathfrak{M}$  of all  $m$ -terminal  $n$ -mesh networks. Let  $\mathfrak{N}(A)$  be an arbitrary network in  $\mathfrak{M}$  having  $A$  for a network matrix. For each  $\mathfrak{N}$  in  $\mathfrak{M}$  select a possible network matrix. Let  $\mathfrak{N}$  range over  $\mathfrak{M}$ . Denote the totality of network matrices so found by  $\mathfrak{A}$ . With each  $\mathfrak{N}(A)$  associate the set  $\mathfrak{L}$  of all networks  $\mathfrak{N}(B)$  whose matrices  $B$  are congruent to  $A$ ,

$$(7.4) \quad B = P'AP,$$

where  $P$  is restricted to the real field. Let  $\mathfrak{P}$  denote the set of all  $P$ 's which satisfy the above requirements.

Next, with  $\mathfrak{N}(A)$  arbitrary, let  $\mathfrak{N}(A_k)$  and  $\mathfrak{N}(B_k)$  denote the networks, having matrices  $A_k$  and  $B_k$ , respectively, obtained by opening the mesh  $k$  of  $\mathfrak{N}(A)$  and  $\mathfrak{N}(B)$ , respectively. We select a maximal subset  $\mathfrak{P}_c$  of  $\mathfrak{P}$  such that for all  $\mathfrak{N}(A)$  of  $\mathfrak{M}$  (that is, for all  $A$  of  $\mathfrak{A}$ ), and for all  $P$  of  $\mathfrak{P}_c$ ,  $B_k = P'_k A_k P_k$ , for  $k = 1, \dots, m$ . In other words, we restrict  $P$  to a set  $\mathfrak{P}_c$  for which  $A_1, A_2, \dots, A_m$  are circavariant matrices of  $A$ . We denote by  $\mathfrak{L}_c$  the subset of  $\mathfrak{L}$  whose matrices  $B, B_1, B_2, \dots, B_m$  are thus related to  $A, A_1, A_2, \dots, A_m$ .

By Theorem 3.2  $P$  must be  $m$ -affine. From Theorems 3.3 and 3.4 we know that  $A_v^*$ , ( $u, v = 1, \dots, m$ ), are also circavariant.

The characteristic coefficient matrices for  $\mathfrak{N}(B)$  and  $\mathfrak{N}(A)$  are

$$(7.5) \quad Y^{(B)} = (Y_{rs}^{(B)}), \quad Y^{(A)} = (Y_{rs}^{(A)}),$$

where

$$Y_{rs}^{(B)} = (-1)^{r+s} \frac{d(B_s^r)}{d(B)}, \quad Y_{rs}^{(A)} = (-1)^{r+s} \frac{d(A_s^r)}{d(A)} \quad (r, s = 1, \dots, m).$$

If  $P$  is  $m$ -affine, by Theorem 4.2 we know that the determinants in  $Y_{rs}^{(B)}$  are circavariants of the set  $\mathfrak{L}$  with respect to  $A_s^r$ . In fact, the admittances

$$Y_{rs}^{(B)} = (-1)^{r+s} \frac{d(P_s^r) \cdot d(A_s^r) \cdot d(P_r)}{d(P) \cdot d(A) \cdot d(P)} = p_{rr}^{-1} y_{rs}^{(A)} p_{ss}^{-1} \quad (r, s = 1, \dots, m)$$

are all relative circavariants. Evidently,

$$(7.6) \quad Y^{(A)} = D'Y^{(B)}D,$$

where

$$D \equiv [p_{11}, \dots, p_{mm}].$$

This shows that every network  $\mathcal{N}(B)$  of  $\mathcal{L}_c$  is circa-equivalent to  $\mathcal{N}(A)$ .

*Case  $m=2$ .* Those networks of  $\mathcal{L}_c$  for which  $p_{rr}=1$  are absolutely equivalent to  $\mathcal{N}(A)$  with respect to terminal pair  $r$ ; those for which  $p_{11}=p_{22} \neq 1$  are relatively equivalent; and those for which  $p_{11}=p_{22}=1$  are absolutely equivalent. By selecting  $P$  so that  $p_{11}p_{22}=1$ ,  $p_{11} \neq 1$ , the transfer admittances of the corresponding  $\mathcal{L}_c$  will all be absolutely circavariant, though the driving-point admittances are only relative circavariants.

If we select a subset  $\mathcal{L}'$  of  $\mathcal{L}$  so that  $A_1^2$  is circavariant, then  $Y_{12}^{(A)}$ , the transfer admittance between meshes 1 and 2, will be a relative circavariant. The requirement that  $A_1^2$  be circavariant makes  $P$  2-affine so that  $\mathcal{L}'$  is  $\mathcal{L}_c$ , and  $Y_{11}^{(A)}$  and  $Y_{22}^{(A)}$  are also relative circavariants.

If we select a subset  $\mathcal{L}_d$  of  $\mathcal{L}$  for which  $A_1$  is circavariant, then the driving-point admittance  $Y_{11}^{(A)}$  is relatively circavariant, but  $Y_{12}^{(A)}$  and  $Y_{22}^{(A)}$  are not necessarily so.

A subset  $\mathcal{L}_e$  of  $\mathcal{L}$  for which  $A_1$  and  $A_2$  are circavariant makes the admittances  $Y_{11}^{(A)}$  and  $Y_{22}^{(A)}$  relative circavariants, with  $P$  2-affine;  $A_1^2$  is then circavariant, so that the admittance  $Y_{12}^{(A)}$  is also relatively circavariant.

**Further results.** More generally, in the case of  $m$ -terminal pair networks, if  $A_{u_1}^{a_1}, A_{u_2}^{a_2}, \dots, A_{u_n}^{a_n}$  be circavariant, then the admittances  $Y_{u_1 a_1}^{(A)}, Y_{u_2 a_2}^{(A)}, \dots, Y_{u_n a_n}^{(A)}$  are relatively circavariant.

The general theory of circa-equivalent networks initiated herein will be developed in greater detail at a later time. It should be noted that the special case when  $D$  is the identity matrix yields the usual theory of (absolutely) equivalent networks.

CASE SCHOOL OF APPLIED SCIENCE,  
CLEVELAND, OHIO

# THE POSITION OF THE RADICAL IN AN ALGEBRA

BY  
MARSHALL HALL

**1. Introduction.** The papers of Brauer, Nesbitt, and Nakayama<sup>(1)</sup> have given a great deal of information about algebras with a radical. These cover wide and interesting, but nevertheless special, classes of algebras. This paper gives a new approach to the study of the most general class of algebras with a radical, starting from the fundamental theorem that every linear associative algebra is uniquely decomposable as the direct sum of a semisimple algebra and an algebra bound to its radical. Here an algebra is said to be bound to its radical (for short: a bound algebra) if the two-sided annihilators of the radical are contained in the radical. In the light of this result, further investigations on algebras with a radical may be confined to bound algebras. A bound algebra is largely determined by its radical. In particular (Theorem 3.8) if the radical is of order  $s$ , the bound algebra is at most of order  $s^2 + s + 1$ .

A combination of the right and left representations of a bound algebra on its radical yields a faithful representation of the algebra modulo the two-sided annihilator of the radical. To obtain a faithful representation of the bound algebra itself (Theorem 3.2) we must adjoin to this representation a system of "remnants" comparable to the factor sets used in the extension of groups or in the theory of normal simple algebras.

A bound algebra is composed (Theorem 3.6) of its radical combined with three orthogonal algebras of which two are semisimple. The third has a unit and is "doubly represented."

The final section of this paper is concerned with the problem of constructing all algebras with a given radical. Some examples are given to illustrate different aspects of this problem.

**2. Decomposition of algebras.** Let  $\mathfrak{A}$  be an arbitrary linear associative algebra and let  $\mathfrak{R}$  be its radical.

Presented to the Society, April 7, 1939; received by the editors March 7, 1940. This paper was received by the editors of the *Annals of Mathematics* June 19, 1939, accepted by them, and later transferred to these *Transactions*.

(<sup>1</sup>) R. Brauer, C. Nesbitt, *On the regular representations of algebras*, *Proceedings of the National Academy of Sciences*, vol. 23 (1937); *On the Modular Representation of Groups of Finite Order*, University of Toronto Studies, Mathematical Series, vol. 4, 1937.

T. Nakayama, C. Nesbitt, *Note on symmetric algebras*, *Annals of Mathematics*, (2), vol. 39 (1938).

C. Nesbitt, *On the regular representations of algebras*, *Annals of Mathematics*, (2), vol. 39 (1938).

T. Nakayama, *Some studies on regular representations, induced representations, and modular representations*, *Annals of Mathematics*, (2), vol. 39 (1938); *On Frobeniusean algebras I*, *ibid.*, vol. 40 (1939); *On Frobeniusean algebras II*, to appear shortly; *On the structure of symmetric algebras and Galois moduli over modular fields*, to appear shortly.

**THEOREM 2.1.** *If  $\mathfrak{A}$  is a left ideal of  $\mathfrak{A}$ , there is an idempotent  $e$  such that  $\mathfrak{a} = (e)_l + \mathfrak{r}_1$  where  $\mathfrak{r}_1 \subset \mathfrak{R}$  and  $\mathfrak{r}_1 e = 0$ . If  $\mathfrak{b}$  is a right ideal, there is an idempotent  $f$  such that  $\mathfrak{b} = (f)_r + \mathfrak{r}_2$  where  $\mathfrak{r}_2 \subset \mathfrak{R}$  and  $\mathfrak{r}_2 f = 0$ . If  $\mathfrak{c}$  is a two-sided ideal, there is an idempotent  $g$  such that  $\mathfrak{c} = (g)_l + \mathfrak{r}_3 = (g)_r + \mathfrak{r}_4$  where  $\mathfrak{r}_3, \mathfrak{r}_4 \subset \mathfrak{R}$  and  $\mathfrak{r}_3 g = \mathfrak{r}_4 = 0$ .*

For the right or left ideals this theorem has been proved by the author<sup>(2)</sup>. A two-sided ideal  $\mathfrak{c}$  may be considered as both a left ideal and a right ideal:

$$(2.1) \quad \begin{aligned} \mathfrak{c} &= (e)_l + \mathfrak{r}_1, & \mathfrak{r}_1 e &= 0, & \mathfrak{r}_1 &\subset \mathfrak{R}, \\ \mathfrak{c} &= (f)_r + \mathfrak{r}_2, & \mathfrak{r}_2 f &= 0, & \mathfrak{r}_2 &\subset \mathfrak{R}. \end{aligned}$$

Here  $f = we + \mathfrak{r}_1$  where  $\mathfrak{r}_1 e = 0$  and  $e = fu + \mathfrak{r}_2$  where  $\mathfrak{r}_2 f = 0$ . Hence  $fe = we = fu$  and  $f - e = \mathfrak{r}_1 - \mathfrak{r}_2 \in \mathfrak{R}$ . If  $x \in (f)_r + \mathfrak{r}_2 = \mathfrak{c}$ ,  $x = fv + s$ ,  $s \in (\mathfrak{R} \cap \mathfrak{c})$ . Hence  $x = (e + \mathfrak{r})v + s = ev + \mathfrak{r}v + s$  where  $\mathfrak{r}v + s \in (\mathfrak{R} \cap \mathfrak{c})$ , and so  $x \in (e)_r \cup (\mathfrak{R} \cap \mathfrak{c})$ . But since  $e \in \mathfrak{c}$ ,  $(e)_r \subset \mathfrak{c}$ , whence  $\mathfrak{c} \subset (e)_r \cup (\mathfrak{R} \cap \mathfrak{c}) \subset \mathfrak{c}$  or  $\mathfrak{c} = (e)_r \cup (\mathfrak{R} \cup \mathfrak{c})$ . But  $(\mathfrak{R} \cap \mathfrak{c}) = e(\mathfrak{R} \cap \mathfrak{c}) + \mathfrak{r}_3$  where  $\mathfrak{r}_3 = 0$ ,  $\mathfrak{r}_3 \subset \mathfrak{R}$ . Hence  $\mathfrak{c} = (e)_r \cup (e(\mathfrak{R} \cap \mathfrak{c}) + \mathfrak{r}_3) = (e)_r + \mathfrak{r}_3$ . Thus in the two representations of (2.1) we may assume without loss of generality that the idempotents  $e$  and  $f$  are the same, which is the statement of the theorem.

**DEFINITION.** *An algebra  $\mathfrak{A}$  is said to be bound to its radical  $\mathfrak{R}$  (briefly:  $\mathfrak{A}$  is a bound algebra) if for  $c \in \mathfrak{A}$ ,  $c\mathfrak{R} = \mathfrak{R}c = 0$  implies that  $c \in \mathfrak{R}$ .*

**THEOREM 2.2.** *Any linear associative algebra is uniquely decomposable as the direct sum of a semisimple algebra and a bound algebra.*

Let  $\mathfrak{A}$  be a linear associative algebra, and let  $\mathfrak{R}$  be its radical. The two-sided annihilators of  $\mathfrak{R}$ , elements  $c$  such that  $c\mathfrak{R} = \mathfrak{R}c = 0$ , form a two-sided ideal  $\mathfrak{R}'$ . By Theorem 2.1 there is an idempotent  $g$  such that  $\mathfrak{R}' = (g)_l + \mathfrak{r}_3 = (g)_r + \mathfrak{r}_4$  where  $\mathfrak{r}_3, \mathfrak{r}_4 \subset \mathfrak{R}$ ,  $\mathfrak{r}_3 g = 0$ ,  $\mathfrak{r}_4 g = 0$ . If

$$(2.2) \quad \mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \mathfrak{A}_3 + \mathfrak{A}_4$$

is the two-sided Pierce decomposition of  $\mathfrak{A}$  with respect to  $g$ , we have, for  $a_i \in \mathfrak{A}_i$ ,

$$(2.3) \quad \begin{aligned} ga_1 &= a_1, & ga_2 &= a_2, & ga_3 &= 0, & ga_4 &= 0, \\ a_1g &= a_1, & a_2g &= 0, & a_3g &= a_3, & a_4g &= 0. \end{aligned}$$

Consider any  $a_2$ .  $\mathfrak{R}a_2 = \mathfrak{R}ga_2 = 0 \cdot a_2 = 0$ . Also  $a_2\mathfrak{R} = ga_2\mathfrak{R} = g(a_2\mathfrak{R}) \subset g\mathfrak{R} = 0$ . Hence  $a_2 \in \mathfrak{R}'$ , and so  $a_2 = wg + \mathfrak{r}$  where  $\mathfrak{r} \in \mathfrak{R}$ ,  $\mathfrak{r}g = 0$ . Here  $0 = a_2g = wg$  and  $a_2 = \mathfrak{r} \in \mathfrak{R}$ . Hence  $a_2 = ga_2 = \mathfrak{r}g \in g\mathfrak{R} = 0$  and  $a_2 = 0$ . Hence  $\mathfrak{A}_2 = 0$ . Similarly  $\mathfrak{A}_3 = 0$ . Hence (2.2) reduces to

<sup>(2)</sup> *A type of algebraic closure*, Annals of Mathematics, (2), vol. 40 (1939), Theorem 6.1. It is evident that the use of a unit in this theorem is not essential.

$$(2.4) \quad \mathfrak{A} = \mathfrak{A}_1 \oplus \mathfrak{A}_4,$$

where the sum is direct since in consequence of the relations (2.3)  $\mathfrak{A}_1\mathfrak{A}_4 = \mathfrak{A}_4\mathfrak{A}_1 = 0$ . Since also  $\mathfrak{A}_1^2 \subset \mathfrak{A}_1$ ,  $\mathfrak{A}_4^2 \subset \mathfrak{A}_4$ ,  $\mathfrak{A}_1$  and  $\mathfrak{A}_4$  are subalgebras of  $\mathfrak{A}$ . Moreover the elements annihilated by  $g$  on both sides are in  $\mathfrak{A}_4$ , and so  $\mathfrak{R} \subset \mathfrak{A}_4$ . If  $\mathfrak{A}_1$  had a radical it would be part of the radical of  $\mathfrak{A}$ , which is in  $\mathfrak{A}_4$ . Hence  $\mathfrak{A}_1$  has no radical, and is semisimple. Moreover  $\mathfrak{R}$  is the radical of  $\mathfrak{A}_4$ . Suppose  $a_4 \in \mathfrak{A}_4$  is a two-sided annihilator of  $\mathfrak{R}$ . Then  $a_4 \in (g)_l + r_s$  and  $a_4 = wg + r_s$  where  $r_sg = 0$ . But  $0 = a_4g = wg$ , and so  $a_4 = r_s \in \mathfrak{R}$ . Hence the two-sided annihilators of  $\mathfrak{R}$  in  $\mathfrak{A}_4$  are in  $\mathfrak{R}$ , and  $\mathfrak{A}_4$  is bound to its radical  $\mathfrak{R}$ .  $\mathfrak{A}$  is the direct sum of the semi-simple algebra  $\mathfrak{A}_1$  and the bound algebra  $\mathfrak{A}_4$ .

Now suppose

$$(2.5) \quad \mathfrak{A} = \mathfrak{B} \oplus \mathfrak{C}$$

is any decomposition of  $\mathfrak{A}$  as the direct sum of a semi-simple algebra  $\mathfrak{B}$  and a bound algebra  $\mathfrak{C}$ . The unit  $h$  of  $\mathfrak{B}$  is a two-sided annihilator of  $\mathfrak{C}$  and a fortiori of  $\mathfrak{R} \subset \mathfrak{C}$ . Hence, under the decomposition (2.4) of  $\mathfrak{A}$ ,  $h = h_1 + h_4$  and each of  $h_1$ ,  $h_4$  is a two-sided annihilator of  $\mathfrak{R}$ . But  $h = h^2 = h_1^2 + h_4^2$ , whence  $h_4^2 = h_4$ . As  $\mathfrak{A}_4$  is a bound algebra,  $h_4 \in \mathfrak{R}$ . An idempotent can be in the radical only if it is zero, and so  $h_4 = 0$ ,  $h = h_1 \in \mathfrak{A}_1$  and  $h = gh = hg$ . A similar argument shows that  $g = hg = gh$ , starting from the decomposition of  $g$  in (2.5). Combining results, we obtain  $g = h$ . Evidently (2.5) as a direct sum is the two-sided Pierce decomposition of  $\mathfrak{A}$  with respect to  $h$  and consequently must be identical with (2.4). This proves the uniqueness part of the theorem.

**3. Representation and properties of bound algebras.** Let  $\mathfrak{A}$  be a bound algebra over a field  $K$  and  $x_1, \dots, x_s$  a basis of its radical  $\mathfrak{R}$ . Then if  $c$  is an arbitrary element of  $\mathfrak{A}$ ,

$$(3.1) \quad \begin{aligned} x_i c &= \sum_{j=1}^s a_{ij} x_j, & i &= 1, \dots, s, \\ c x_i &= \sum_{j=1}^s b_{ij} x_j, & i &= 1, \dots, s, \end{aligned}$$

and we have the right representation of  $\mathfrak{A}$  on  $\mathfrak{R}$ :

$$(3.2) \quad c \rightarrow (a_{ij}) = R(c)$$

and the left representation<sup>(\*)</sup> of  $\mathfrak{A}$  on  $\mathfrak{R}$ :

$$(3.3) \quad c \rightarrow (b_{ij}) = L(c).$$

Now (3.2) is a faithful representation of  $\mathfrak{A}/\mathfrak{R}$  since the  $c$ 's mapped onto zero are the right annihilators of  $\mathfrak{R}$ . Similarly (3.3) is a faithful representation

(\*) For the left representation  $cd \rightarrow L(d)L(c)$ . Ordinarily the transpose  $L(c)^T$  is used to preserve the order of multiplication. But in this paper it seems desirable to leave the representation in this form.



of  $\mathfrak{A}/\mathfrak{N}^t$ . (Note that  $\mathfrak{N}^r$  and  $\mathfrak{N}^t$  are both two-sided ideals.) A combination of (3.2) and (3.3) is better than either one separately. If we put

$$(3.4) \quad c \rightarrow [R(c), L(c)],$$

we have the rules of combination

$$(3.5) \quad \begin{aligned} c_1 + c_2 &\rightarrow [R(c_1) + R(c_2), L(c_1) + L(c_2)], \\ c_1 c_2 &\rightarrow [R(c_1)R(c_2), L(c_2)L(c_1)], \\ kc &\rightarrow [kR(c), kL(c)], \end{aligned} \quad k \in K.$$

Here (3.4) subject to the combinatory rules (3.5) is a faithful representation of  $\mathfrak{A}$  modulo  $\mathfrak{N}^t$ .

**THEOREM 3.1.** *In the representation (3.4) every matrix  $R(c_1)$  permutes with every matrix  $L(c_2)$ .*

This well known theorem on representations is an immediate consequence of the associative law  $c_2(\mathfrak{N}c_1) = (c_2\mathfrak{N})c_1$ .

To obtain a faithful representation of  $\mathfrak{A}$  we must extend the representation (3.4) by the adjunction of  $\mathfrak{N}^t$ . Let  $z_1, \dots, z_q$  be a basis of  $\mathfrak{N}^t$  and extend this to a basis of  $\mathfrak{N}$ ,  $u_1, \dots, u_p, z_1, \dots, z_q$ , where  $p+q=s$  and finally to a basis of  $\mathfrak{A}$ ,  $u_1, \dots, u_p, u_{p+1}, \dots, u_m, z_1, \dots, z_q$ . If, in the homomorphism  $\mathfrak{A} \rightarrow \mathfrak{A}/\mathfrak{N}^t$ ,  $u_i \rightarrow u_i$ , then  $u_1, \dots, u_m$  form a basis of  $\mathfrak{A}/\mathfrak{N}^t$ . We shall call  $u_i$  the representative of its class in  $\mathfrak{A}$  modulo  $\mathfrak{N}^t$ . To extend this concept of representative, suppose an arbitrary  $c \in \mathfrak{A}$  is given by

$$(3.6) \quad c = \sum_{j=1}^m c_j u_j + \sum_{k=1}^q r_k z_k.$$

Then the mapping  $\mathfrak{A} \rightarrow \mathfrak{A}/\mathfrak{N}^t$ , which we may suppose given by (3.4), maps  $c$  onto an element  $\gamma = \sum_{j=1}^m c_j u_j$ . If we now write

$$(3.7) \quad \tilde{\gamma} = \sum_{j=1}^m c_j u_j,$$

we call  $\tilde{\gamma}$  the representative of the class of  $\mathfrak{A}$  modulo  $\mathfrak{N}^t$  mapped onto  $\gamma$ . Hence any  $c$  of  $\mathfrak{A}$  is expressible uniquely as

$$(3.8) \quad c = \tilde{\gamma} + r,$$

where  $c \rightarrow \gamma$  by the homomorphism (3.4) and  $r \in \mathfrak{N}^t$ .

Let us now suppose the basis of  $\mathfrak{N}$  used in (3.4) is chosen to be  $u_1, \dots, u_p, z_1, \dots, z_q$ . Since  $\mathfrak{N}^t$  is a two-sided ideal, the matrices  $R(c)$  and  $L(c)$  must be of the form

$$(3.9) \quad R(c) = \begin{pmatrix} R_{11}(c) & R_{12}(c) \\ 0 & R_{22}(c) \end{pmatrix}, \quad L(c) = \begin{pmatrix} L_{11}(c) & L_{12}(c) \\ 0 & L_{22}(c) \end{pmatrix},$$

where  $R_{22}(c)$  and  $L_{22}(c)$  are  $q$  by  $q$  matrices giving the right and left transformations induced on  $\mathfrak{R}^t$  by  $c$ . If

$$(3.10.1) \quad r = \sum_{k=1}^q r_k s_k$$

is any element of  $\mathfrak{R}^t$ , then

$$(3.10.2) \quad rc = \sum_{k=1}^q t_k s_k, \quad cr = \sum_{k=1}^q s_k s_k,$$

where

$$(3.10.3) \quad \begin{pmatrix} t_1 \\ \vdots \\ t_q \end{pmatrix} = R_{22}(c) \begin{pmatrix} r_1 \\ \vdots \\ r_q \end{pmatrix}, \quad \begin{pmatrix} s_1 \\ \vdots \\ s_q \end{pmatrix} = L_{22}(c) \begin{pmatrix} r_1 \\ \vdots \\ r_q \end{pmatrix}.$$

**THEOREM 3.2. REPRESENTATION OF BOUND ALGEBRAS.** Let  $\mathfrak{A}$  be an algebra bound to its radical  $\mathfrak{R}$ . Then a faithful representation of  $\mathfrak{A}$  is given by

$$(3.11) \quad c \rightleftharpoons [R(c), L(c), W(c)],$$

where  $R(c)$ ,  $L(c)$  are given by (3.1)–(3.3) and are of the type (3.9) and  $W(c) = (r_1 \cdots r_q)'$  (the prime indicating the transpose of the vector) is determined by  $r = \sum_{k=1}^q r_k s_k$  in (3.6). The rules of combination in (3.11) are given by

$$(3.12) \quad \begin{aligned} kc &\rightleftharpoons [kR(c), kL(c), kW(c)], & k \in K, \\ c_1 + c_2 &\rightleftharpoons [R(c_1) + R(c_2), L(c_1) + L(c_2), W(c_1) + W(c_2)], \\ c_1 c_2 &\rightleftharpoons [R(c_1)L(c_2), L(c_2)L(c_1), R_{22}(c_2)W(c_1) + L_{22}(c_1)W(c_2) + \{\gamma_1, \gamma_2\}]. \end{aligned}$$

Here  $\{\gamma_1, \gamma_2\} = (d_1, \dots, d_q)'$  is determined by

$$(3.13) \quad \tilde{\gamma}_1 \tilde{\gamma}_2 = \overline{\gamma_1 \gamma_2} + r(\gamma_1, \gamma_2); \quad r(\gamma_1, \gamma_2) = \sum_{k=1}^q d_k s_k \in \mathfrak{R}^t.$$

**Proof.** It is easily seen that (3.11) yields a one-to-one correspondence between the elements of  $\mathfrak{A}$  and the symbols  $[R(c), L(c), W(c)]$ . For (3.8) expresses  $c$  uniquely as the sum  $\tilde{\gamma} + r$  and  $\tilde{\gamma} \rightleftharpoons \gamma$ ,  $\gamma \rightleftharpoons [R(c), L(c)]$ ,  $r \rightleftharpoons W(c)$ . It remains to show that the rules of combination (3.12) are in accord with this correspondence. For the sum and scalar product this is evident. For the product

$$(3.14) \quad \begin{aligned} c_1 &= \tilde{\gamma}_1 + r_1, & c_2 &= \tilde{\gamma}_2 + r_2, \\ c_1 c_2 &= \tilde{\gamma}_1 \tilde{\gamma}_2 + \tilde{\gamma}_1 r_2 + r_1 \tilde{\gamma}_2 = \overline{\gamma_1 \gamma_2} + \tilde{\gamma}_1 r_2 + r_1 \tilde{\gamma}_2 + r(\gamma_1, \gamma_2), \end{aligned}$$

where  $\overline{\gamma_1 \gamma_2} \rightleftharpoons [R(c_1)R(c_2), L(c_2)L(c_1)]$  from (3.5) and  $\tilde{\gamma}_1 r_2 \rightleftharpoons L_{22}(c_1)W(c_2)$ ,  $r_1 \tilde{\gamma}_2 \rightleftharpoons R_{22}(c_2)W(c_1)$  by (3.10.3) and  $r(\gamma_1, \gamma_2) \rightleftharpoons \{\gamma_1, \gamma_2\}$  by (3.13). The element

$r(\gamma_1, \gamma_2)$  will be called the *remnant* of the product  $\tilde{\gamma}_1, \tilde{\gamma}_2$ . Here  $r_1 r_2 = 0$  since  $(\mathfrak{R}')^2 = 0$  because  $(\mathfrak{R}')^2 \subset \mathfrak{R}'\mathfrak{R} = 0$ .

**THEOREM 3.3.** *The remnants  $r(x, y)$  of (3.13) satisfy the following relations:*

$$\begin{aligned}
 (3.15) \quad & r(x + y, z) = r(x, z) + r(y, z), \\
 & r(x, y + z) = r(x, y) + r(x, z), \\
 & r(kx, y) = r(x, ky) = kr(x, y), \quad k \in K, \\
 & \bar{x}y r(u, v) = \overline{xy} r(u, v), \\
 & r(u, v) \bar{x}y = r(u, v) \overline{xy}, \\
 & \bar{x}r(y, z) + r(x, yz) = r(xy, z) + r(x, y)\bar{z}.
 \end{aligned}$$

From the definition of the representative  $\tilde{y}$  in (3.7) it follows immediately that

$$(3.16) \quad \overline{x + y} = \bar{x} + \bar{y}, \quad \overline{kx} = k\bar{x}, \quad k \in K;$$

whence the first three relations are easily derived. For the fourth relation,  $\bar{x}y r(u, v) = (\overline{xy} + r(x, y))r(u, v) = \overline{xy}r(u, v)$  since  $(\mathfrak{R}')^2 = 0$ . The fifth relation may be derived in the same way. The last and perhaps the most important relation is obtained by multiplying out  $\bar{x}(yz) = (\bar{x}y)\bar{z}$ . In constructing a bound algebra with a given radical it is this last relation which is most difficult to satisfy.

If the representatives  $\bar{u}_1, \dots, \bar{u}_m$  are replaced by representatives  $\bar{\bar{u}}_1, \dots, \bar{\bar{u}}_m$  in the same classes of  $\mathfrak{A}$  modulo  $\mathfrak{R}'$ , then

$$\begin{aligned}
 (3.17) \quad & \bar{\bar{u}}_i = \bar{u}_i + \alpha(u_i), \quad i = 1, \dots, m, \\
 & \bar{\bar{x}} = \bar{x} + \alpha(x),
 \end{aligned}$$

where  $\alpha(x)$  satisfies the linearity conditions

$$(3.18) \quad \alpha(x + y) = \alpha(x) + \alpha(y), \quad \alpha(kx) = k\alpha(x), \quad k \in K.$$

**THEOREM 3.4.** *If the representatives of the classes of  $\mathfrak{A}$  modulo  $\mathfrak{R}'$  are changed by the rule (3.17), then the remnants  $r(x, y)$  are changed by the rule*

$$(3.19) \quad r'(x, y) = r(x, y) + \bar{x}\alpha(y) + \alpha(x)\bar{y} - \alpha(xy).$$

**Proof.**  $\bar{\bar{x}}\bar{\bar{y}} = (\bar{x} + \alpha(x))(\bar{y} + \alpha(y))$  or  $\bar{\bar{x}}\bar{\bar{y}} + r'(x, y) = \bar{x}\bar{y} + \bar{x}\alpha(y) + \alpha(x)\bar{y}$  or  $\bar{\bar{x}}\bar{\bar{y}} + \alpha(xy) + r'(x, y) = \bar{x}\bar{y} + r(x, y) + \bar{x}\alpha(y) + \alpha(x)\bar{y}$ , whence (3.19) follows.

**THEOREM 3.5.** *If  $e_1, \dots, e_n$  are orthogonal idempotents in  $\mathfrak{A}/\mathfrak{R}'$ , then the representatives  $\bar{e}_1, \dots, \bar{e}_n$  may be chosen as orthogonal idempotents.*

The proof of this theorem exactly parallels the proof of Theorem 1 on page 16 of Deuring's *Algebren*.

**THEOREM 3.6.** *In a bound algebra  $\mathfrak{A}$  there are subalgebras  $\mathfrak{A}_1, \mathfrak{A}_r, \mathfrak{A}_d$  such that  $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_r + (\mathfrak{A}_d \cup \mathfrak{N})$  with the following properties:*

1.  $\mathfrak{A}_1, \mathfrak{A}_r, \mathfrak{A}_d$  have units and are orthogonal.
2.  $\mathfrak{A}_1$  and  $\mathfrak{A}_r$  are semisimple and  $\mathfrak{N}\mathfrak{A}_1 = 0, \mathfrak{A}_r\mathfrak{N} = 0$ .
3.  $\mathfrak{A}/\mathfrak{N}$  is the direct sum of three semisimple algebras isomorphic to  $\mathfrak{A}_1, \mathfrak{A}_r, \mathfrak{A}_d/(\mathfrak{A}_d \cap \mathfrak{N})$ .

**Proof.**  $\mathfrak{N}^r$  and  $\mathfrak{N}^i$  are both two-sided ideals and  $\mathfrak{A}$  itself may be considered a two-sided ideal, whence by Theorem 2.1

$$(3.20) \quad \begin{aligned} \mathfrak{N}^r &= (e_1)_r + r_1, & \mathfrak{N}^r &= (e_1)_i + r_2, & \mathfrak{N}^i &= (e_2)_r + r_3, \\ \mathfrak{N}^i &= (e_2)_i + r_4, & \mathfrak{A} &= (e)_r + r_5, & \mathfrak{A} &= (e)_i + r_6, \end{aligned}$$

with relations on the  $r$ 's as given in Theorem 2.1.

Put  $e_3 = e - e_1 - e_2$ . Then it may be shown that  $e_i e_j, i \neq j$ , is a two-sided annihilator of  $\mathfrak{N}$  and that consequently the images of  $e_1, e_2$ , and  $e_3$  in  $\mathfrak{A}/\mathfrak{N}^i$  are orthogonal idempotents. By Theorem 3.5 there exist orthogonal idempotents  $e'_1, e'_2$ , and  $e'_3$  in the classes of  $e_1, e_2$ , and  $e_3$  modulo  $\mathfrak{N}^i$ . Just as in the proof of Theorem 2.1 it may be shown that  $e'_1, e'_2$ , and  $e' = e'_1 + e'_2 + e'_3$  may be used in the representation of the ideals in (3.20). Stated formally:

**LEMMA.** *Without loss of generality it may be assumed that the idempotents of (3.20) satisfy the following relations:*

$$(3.21) \quad ee_1 = e_1e = e_1, \quad ee_2 = e_2e = e_2, \quad e_1e_2 = e_2e_1 = 0.$$

Now put  $\mathfrak{A}_1 = e_1\mathfrak{A}e_1$ ,  $\mathfrak{A}_r = e_2\mathfrak{A}e_2$ , and  $\mathfrak{A}_d = e_3\mathfrak{A}e_3$ . Here  $e_1, e_2$ , and  $e_3$  are the units of  $\mathfrak{A}_1, \mathfrak{A}_r$ , and  $\mathfrak{A}_d$  respectively, and the orthogonality of these algebras is an immediate consequence of the orthogonality of these idempotents. This is the first property mentioned in the theorem.

To prove the decomposition

$$(3.22) \quad \mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_r + (\mathfrak{A}_d \cup \mathfrak{N}),$$

take any  $x$  of  $\mathfrak{A}$ . From the relations (3.20)  $x = ew + t$  where  $t \in \mathfrak{N}$ ,  $et = 0$ . Here  $ex = ew$ ,  $x = ex + t$ . Also  $ex = ue + s$  with  $s \in \mathfrak{N}$ ,  $se = 0$  and so  $exe = ue$ ,  $ex = exe + s$ ,  $x = exe + s + t = exe + \rho$  with  $\rho \in \mathfrak{N}$ . Hence  $x = (e_1 + e_2 + e_3)x(e_1 + e_2 + e_3) + \rho = e_1xe_1 + e_2xe_2 + e_3xe_3 + \rho^* = x_i + x_r + (x_d + \rho^*)$ . Here  $\rho^* = \sum_{i=1,2,3} e_i x e_i + \rho \in \mathfrak{N}$  since  $e_i x e_i$  is in the radical. For  $e_1 x e_1 \in \mathfrak{N}^r = (e_1)_i + r_2$ ,  $e_1 x e_1 = ue_1 + t$  where  $t \in \mathfrak{N}$ ,  $te_1 = 0$ . Hence  $0 = e_1 x e_1 e_1 = ue_1$  and  $e_1 x e_1 = t \in \mathfrak{N}$ . A similar argument holds for all the  $e_i x e_i, j \neq i$ . To show that the sum  $x = x_i + x_r + (x_d + \rho^*)$  is unique it is enough to show that

$$(3.23) \quad 0 = x_i + x_r + (x_d + \rho^*)$$

implies that  $x_i, x_r$ , and  $x_d + \rho^*$  all vanish. For  $0 = x_i e_1 + x_r e_1 + (x_d + \rho^*) e_1$  and here  $x_r e_1 = x_r e_2 e_1 = 0$ ,  $x_d e_1 = x_d e_3 e_1 = 0$ , and since  $\rho^* \in \mathfrak{N}$ ,  $\rho^* e_1 = 0$ . Hence  $x_i = x_i e_1$

$= 0$ . Then  $0 = e_2 x_r + e_2(x_d + \rho^*)$  and  $e_2 x_d = 0$ ,  $e_2 \rho^* = 0$ , and so  $x_r = e_2 x_r = 0$ . Now as  $x_i = 0$ ,  $x_r = 0$ , then from (3.23)  $x_d + \rho^* = 0$ . This proves the decomposition (3.22).

For the second property, since  $\mathfrak{A}_r = e_2 \mathfrak{A} e_2$  and  $e_2 \mathfrak{R} = 0$  it follows that  $\mathfrak{A}_r \mathfrak{R} = 0$ . Similarly  $\mathfrak{R} \mathfrak{A}_i = 0$ .

It remains to show that  $\mathfrak{A}_i$  and  $\mathfrak{A}_r$  are semisimple. If  $\mathfrak{A}_i$  contained a nilpotent ideal  $\mathfrak{r}$ , then

$$\mathfrak{r} \mathfrak{A} \subset \mathfrak{r} \mathfrak{A}_i + \mathfrak{r} \mathfrak{R} = \mathfrak{r} + \mathfrak{r} \mathfrak{R}$$

by the decomposition (3.22) and the orthogonality of  $\mathfrak{A}_i$ ,  $\mathfrak{A}_r$ , and  $\mathfrak{A}_d$ . Here  $\mathfrak{r} + \mathfrak{r} \mathfrak{R}$  would be a nilpotent ideal in  $\mathfrak{A}$  and hence contained in  $\mathfrak{R}$ , whence  $\mathfrak{r}$  would be contained in  $\mathfrak{R}$ . But  $\mathfrak{A}_i$  cannot contain any elements of  $\mathfrak{R}$  since  $e_2$ , the unit of  $\mathfrak{A}_r$ , is a left annihilator of  $\mathfrak{R}$ . In the same way it may be shown that  $\mathfrak{A}_r$  is semisimple.

Applying the homomorphism  $\mathfrak{A} \rightarrow \mathfrak{A}/\mathfrak{R}$  to the decomposition (3.22) we have

$$(3.24) \quad \mathfrak{A}/\mathfrak{R} = \tilde{\mathfrak{A}}_i \oplus \tilde{\mathfrak{A}}_r \oplus \tilde{\mathfrak{A}}_d,$$

$\tilde{\mathfrak{A}}_i$  being the image of  $\mathfrak{A}_i$ . The sum is direct since the  $e_i$  and a fortiori their images are orthogonal. Since  $\mathfrak{A}_i$  and  $\mathfrak{A}_r$  are semisimple they must be isomorphic to their images. And as  $\mathfrak{R} \rightarrow 0$   $\mathfrak{A}_d \cup \mathfrak{R} \rightarrow \tilde{\mathfrak{A}}_d \cup 0 = \tilde{\mathfrak{A}}_d$ . The radical of  $\mathfrak{A}_d$  is  $\mathfrak{A}_d \cap \mathfrak{R}$  and so  $\tilde{\mathfrak{A}}_d = \mathfrak{A}_d/(\mathfrak{A}_d \cap \mathfrak{R})$ .

**THEOREM 3.7.**  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_r$ ,  $\mathfrak{R} \rightarrow \mathfrak{A}_i \mathfrak{R}$  are faithful representations of  $\mathfrak{A}_r$  and  $\mathfrak{A}_i$  respectively. Neither of the mappings  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_d$  and  $\mathfrak{R} \rightarrow \mathfrak{A}_d \mathfrak{R}$  maps onto zero any element of  $\mathfrak{A}_d$  not in  $\mathfrak{R}$ . Hence  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_d$  is a faithful representation of  $\mathfrak{A}_d/b_1$  and  $\mathfrak{R} \rightarrow \mathfrak{A}_d \mathfrak{R}$  is a faithful representation of  $\mathfrak{A}_d/b_2^{(*)}$  where  $b_1, b_2 \subset \mathfrak{R}$ ,  $b_1^2 = 0$ ,  $b_2^2 = 0$ .

Suppose the mapping  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_r$  represents some  $z$  of  $\mathfrak{A}_r$  as 0. Then  $\mathfrak{R} z = 0$ , but  $\mathfrak{A}_r \mathfrak{R} = 0$ , and so  $z \mathfrak{R} = 0$ . Hence as a two-sided annihilator of  $\mathfrak{R}$ ,  $z$  belongs to  $\mathfrak{R}$ . But by the preceding theorem  $\mathfrak{A}_r$  contains no elements of  $\mathfrak{R}$ . Hence  $z = 0$  and  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_r$  is a faithful representation. Similarly  $\mathfrak{R} \rightarrow \mathfrak{A}_i \mathfrak{R}$  is a faithful representation.

If  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_d$  represents a  $z \in \mathfrak{A}_d$  by zero, then  $\mathfrak{R} z = 0$ ,  $z \in \mathfrak{R}' = (e_1)_r + \mathfrak{r}_1 = (e_1)_i + \mathfrak{r}_2$ . For any  $w \in \mathfrak{R}'$ ,  $w = e_1 w e_1 + t$  with  $t \in \mathfrak{R}$ ,  $e_1 w e_1 \in \mathfrak{A}_i$ , and so for  $z \in \mathfrak{R}'$ ,  $z \in \mathfrak{A}_d$ ,  $z = t \in \mathfrak{R}$ . Hence the only elements of  $\mathfrak{A}_d$  represented by zero in  $\mathfrak{R} \rightarrow \mathfrak{R} \mathfrak{A}_d$  are elements of  $\mathfrak{R}$ . Those elements mapped onto zero form a two-sided ideal  $b_1$  in  $\mathfrak{R}$ , and since  $\mathfrak{R} b_1 = 0$  a fortiori  $b_1^2 = 0$ . Similarly we may treat the representation  $\mathfrak{R} \rightarrow \mathfrak{A}_d \mathfrak{R}$ .

**THEOREM 3.8.** If a radical  $\mathfrak{R}$  is of order  $s$ , then an algebra  $\mathfrak{A}$  bound to  $\mathfrak{R}$  is at most of order  $s^2 + s + 1$ .

(\*) These two algebras may be different. See Example 2 in §4.



**Proof.** Consider the decomposition (3.22) of  $\mathfrak{A}$  and the representation (3.11) of  $\mathfrak{A}$ . Since neither  $\mathfrak{A}_l$  nor  $\mathfrak{A}_r$  contains any elements of  $\mathfrak{R}'$ , we may choose their elements as representatives of their classes modulo  $\mathfrak{R}'$  in (3.11) and hence

$$(3.25) \quad \begin{aligned} c &\not\equiv [0, L(c), 0], & c &\in \mathfrak{A}_l, \\ c &\not\equiv [R(c), 0, 0], & c &\in \mathfrak{A}_r; \end{aligned}$$

whence  $\mathfrak{A}_l$  may be called the left represented subalgebra,  $\mathfrak{A}_r$  the right represented subalgebra. In the sense of Theorem 3.7  $\mathfrak{A}_d$  may be called the doubly represented subalgebra.

Now we appeal to the theorems on fully reducible matrix algebras as they appear in Weyl's *The Classical Groups, Their Invariants and Representations*, chap. 3. As  $\mathfrak{A}_l$  is semisimple, the representation  $\mathfrak{R} \rightarrow \mathfrak{A}_l \mathfrak{R}$  is fully reducible, the irreducible components corresponding to the simple algebras whose direct sum is  $\mathfrak{A}_l$ . If  $\mathfrak{A}_l = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$  where the  $\mathfrak{A}_i$  are simple, then the representation breaks up into blocks of degrees  $g_1, \dots, g_n, g_{n+1} = s - (g_1 + g_2 + \cdots + g_n)$ , the  $i$ th block containing a certain number of equivalent irreducible representations of  $\mathfrak{A}_i$ , and the last vanishing. The commutator algebra of  $\mathfrak{A}_l$  will break up into blocks  $B_1, \dots, B_n, B_{n+1}$  where  $B_i$  is the commutator algebra of the  $i$ th block,  $i = 1, \dots, n$ , and  $B_{n+1}$  is the complete  $g_{n+1}^2$  matrix algebra. From Weyl, page 93, the orders  $h_i$  and  $h'_i$  of  $\mathfrak{A}_i$  and  $B_i$  respectively satisfy  $h_i h'_i = g_i^2$  for  $i = 1, \dots, n$  and  $h'_{n+1} = g_{n+1}^2$ .

Now every element of  $\mathfrak{A}_r$  and every element of  $\mathfrak{A}_d$  not in  $\mathfrak{R}$  has a proper representation  $c \not\equiv R(c)$  from Theorem 3.7, and by Theorem 3.1  $R(c)$  must be a subalgebra of the commutator algebra of  $L(c)$ . Hence the order of  $\mathfrak{A}$  does not exceed the order of  $\mathfrak{A}_l$  plus the order of the commutator of  $\mathfrak{A}_l$  plus the order of  $\mathfrak{R}$ . Hence the order of  $\mathfrak{A}$  is at most

$$\sum_{i=1}^n (h_i + h'_i) + g_{n+1}^2 + s = \sum_{i=1}^n (h_i + g_i^2/h_i) + g_{n+1}^2 + s.$$

The  $g$ 's and  $h$ 's are positive integers ( $g_{n+1}$  might be zero) and the sum of the  $g$ 's is  $s$ . Here it is very easy to show that

$$\sum_{i=1}^n (h_i + g_i^2/h_i) + g_{n+1}^2 \leq s^2 + s + 1$$

and that equality holds only when  $n=1$ ,  $g_1=s$ , and  $h_1=1$  or  $s^2$ . This proves the theorem. This result is the best possible since we could have an algebra with  $\mathfrak{R}^2=0$ ,  $\mathfrak{A}_l$  the complete  $s^2$  matrix algebra and  $\mathfrak{A}_r$  the scalar algebra of order 1. On the other hand, when  $\mathfrak{R}^2 \neq 0$  the order of  $\mathfrak{A}$  must be less than  $s^2+s+1$  and it should be possible to obtain various improvements on this theorem.

4. **Construction of bound algebras.** In constructing all algebras bound to

a given radical  $\mathfrak{R}$ , we turn to representation (3.11), remembering that from Theorem 3.6 we need determine only  $\mathfrak{A}_r, \mathfrak{A}_l, \mathfrak{A}_d$  separately. When  $\mathfrak{A}_d$  is void, equations (3.25) make the construction of the algebra relatively easy. Any two semisimple matrix algebras which permute with each other and the representations of elements of  $\mathfrak{R}$  may be considered the right and left represented subalgebras of an algebra bound to  $\mathfrak{R}$ . The construction of algebras in which  $\mathfrak{A}_d$  is not void offers many more difficulties. In the first place  $\mathfrak{A}_d$  is not usually semisimple and its right and left representations may be different algebras. Moreover the faithful representation of  $\mathfrak{A}_d$  in (3.11) may involve remnants which must be chosen to satisfy equations (3.15).

**THEOREM 4.1.** *Given a nilpotent algebra  $\mathfrak{R}$  of order  $s$ , let the right representation of  $r \in \mathfrak{R}$  on a basis of  $\mathfrak{R}$  be  $r \rightarrow R(r)$  and the left representation be  $r \rightarrow L(r)$ . If, in the homomorphism  $\mathfrak{R} \rightarrow \mathfrak{R}/\mathfrak{R}'$ ,  $r \rightarrow \rho$ , then  $\rho \mapsto [R(r), L(r)]$  is a faithful representation of  $\mathfrak{R}/\mathfrak{R}'$ . Suppose (1)  $c \mapsto [R(c), L(c)]$  is a right-left representation of an  $s$  by  $s$  matrix algebra  $\mathfrak{A}'$  whose radical is  $\mathfrak{R}/\mathfrak{R}'$  and that every  $R(c_1)$  permutes with every  $L(c_2)$ ; (2) remnants  $r(x, y)$  are chosen from  $\mathfrak{R}'$  for every pair  $x, y$  of elements of  $\mathfrak{A}'$  such that equations (3.15) are satisfied; and (3) the remnants  $r(\rho_i, \rho_j)$  are such that  $\bar{\rho}_i$  may be considered representatives of classes of  $\mathfrak{R}$  modulo  $\mathfrak{R}'$ . Then (3.11) yields a faithful representation of an algebra bound to  $\mathfrak{R}$ , the rules of combination being given by (3.12) and (3.13).*

Theorem 3.2 shows that every bound algebra has a representation (3.11). This theorem shows that conversely the symbols (3.11) define an algebra bound to  $\mathfrak{R}$  providing that certain conditions are satisfied. The proof is direct though a little tedious and will only be sketched here. It must be shown that the rules (3.12) and (3.13) actually define an associative algebra and it is here that we need equations (3.15) and the permutability of  $R(c_1)$  and  $L(c_2)$ . Moreover conditions 1 and 2 assure us that the radical of this algebra is  $\mathfrak{R}$ , and neither more nor less. That  $\mathfrak{A}$  is bound to  $\mathfrak{R}$  is an immediate consequence of the fact that the elements of  $\mathfrak{A}$  are properly represented on  $\mathfrak{R}$  apart from  $\mathfrak{R}' \subset \mathfrak{R}$ .

In practise the following theorem is of use:

**THEOREM 4.2.**  *$c \rightarrow R(c)$  is a faithful representation of  $\mathfrak{A}/\mathfrak{R}'$ , and the radical of  $[R(c)]$  is the right representation of  $\mathfrak{R}$ .  $\mathfrak{A}/\mathfrak{R}'$  modulo its radical is isomorphic to  $\mathfrak{A}_r \oplus \mathfrak{A}_d$ . Similarly the radical of  $[L(c)]$  is the left representation of  $\mathfrak{R}$ , and  $[L(c)]$  modulo its radical is isomorphic to  $\mathfrak{A}_l \oplus \mathfrak{A}_d$ .*

**Proof.** The matrix algebra  $[R(c)]$  is  $\mathfrak{A}/\mathfrak{R}'$  since it is a homomorphic image of  $\mathfrak{A}$  in which the elements mapped onto zero are those of  $\mathfrak{R}'$ . Suppose  $T$  is its radical. The elements of  $\mathfrak{A}$  mapped onto  $T$  and zero form a two-sided ideal  $K$ , and include  $\mathfrak{A}_l$  and  $\mathfrak{R}$ . Since  $T^2 = 0$ ,  $K^2 \subset \mathfrak{R}'$ . From Theorem 2.1  $K = (e)_r + \mathfrak{r}$  with  $\mathfrak{r} \subset \mathfrak{R}$ . Now  $e \in K^2 \subset \mathfrak{R}'$ . Hence  $(e)_r$  is represented by zero and  $T$  is the representation of  $\mathfrak{r}$  alone. Hence  $T$  is the image of  $\mathfrak{R}$ , and the only elements

of  $\mathfrak{A}$  mapped onto  $T$  and zero are  $\mathfrak{A}_1 + \mathfrak{N}$ . From the decomposition (3.22)  $\mathfrak{A}/\mathfrak{N}$  modulo its radical is isomorphic to  $\tilde{\mathfrak{A}}_1 \oplus \tilde{\mathfrak{A}}_2$ . The proof for  $[L(c)]$  is similar.

EXAMPLE 1.  $\mathfrak{N}$  is cyclic:  $\mathfrak{N} = (x, \dots, x^{n-1})$ , where  $x^n = 0$ .

Case 1.  $\mathfrak{A} = \mathfrak{N}$ .

Case 2.  $\mathfrak{A} \neq \mathfrak{N}$  and  $\mathfrak{A} = \mathfrak{N}^i \cup \mathfrak{N}$ . Here  $\mathfrak{A} = (e_2)_r + r$ . Now  $xe_2 = a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}$  and since  $e_2^2 = e_2$ ,  $x^n = 0$ , it follows that  $xe_2 = 0$  or  $x$ . Similarly  $e_2x = 0$  or  $x$ . As  $\mathfrak{A} = \mathfrak{N}^i \cup \mathfrak{N}$ ,  $e_2x = 0$ . Now if also  $xe_2 = 0$ , then  $e_2 \in \mathfrak{N}^i \subset \mathfrak{N}$ , which is impossible. Hence  $xe_2 = x$ . An arbitrary  $y \in \mathfrak{A}$  is of the form  $y = e_2ye_2 + t$  with  $t \in \mathfrak{N}$ . Let  $xe_2ye_2 = b_1x + \dots + b_{n-1}x^{n-1}$ . Then put  $v = e_2ye_2 - b_1e_2$ . Here  $vx = 0$ ,  $xv = b_2x^2 + \dots + b_{n-1}x^{n-1}$ . From this  $(v)_r^n \subset \mathfrak{N}^r$ , and, since  $(v)_r^n \subset \mathfrak{N}^i$ ,  $(v)_r^n \subset \mathfrak{N}^i \subset \mathfrak{N}$ , whence  $(v)_r$  is a nilpotent ideal and  $v \in \mathfrak{N}$ . Hence  $v = e_2ve_2 \in e_2\mathfrak{N}e_2 = 0$  and  $v = 0$ . Hence  $y = b_1e_2 + t$  where  $t \in \mathfrak{N}$ . Here  $\mathfrak{A} = (e_2, x, \dots, x^{n-1})$  with  $e_2x = 0$ ,  $xe_2 = x$ .

Case 3.  $\mathfrak{A} \neq \mathfrak{N}^i \cup \mathfrak{N}$ . Here  $\mathfrak{A} = (e)_i + r$ , and, arguing as above, we conclude  $ex = x$ ,  $xe = 0$  or  $x$  while  $\mathfrak{A} = (e) + \mathfrak{N}^i \cup \mathfrak{N}$ . We distinguish according as  $xe = 0$  or  $x$ , and as  $\mathfrak{N}^i \cup \mathfrak{N} = \mathfrak{N}$  or  $(e_2, \mathfrak{N})$ .

Case 3.1.  $\mathfrak{A} = (e_1, e_2, \mathfrak{N})$ :

$$\begin{aligned} e_1^2 &= e_1, & e_2^2 &= e_2, & e_1e_2 &= e_2e_1 = 0, \\ e_1x &= x, & xe_1 &= 0, & e_2x &= 0, & xe_2 &= x. \end{aligned}$$

Case 3.2.  $\mathfrak{A} = (e_1, \mathfrak{N})$ :

$$e_1^2 = e_1, \quad e_1x = x, \quad xe_1 = 0.$$

Case 3.3.  $\mathfrak{A} = (e, \mathfrak{N})$ :

$$e^2 = e, \quad ex = xe = x.$$

Thus in all five cases and for any radical five similar bound algebras will exist. Let  $\mathfrak{N}$  be any nilpotent algebra and  $e_1$  and  $e_2$  two orthogonal idempotents. For any  $r \in \mathfrak{N}$  let  $e_1r = r$ ,  $re_1 = 0$ ,  $e_2r = 0$ ,  $re_2 = r$ . Then we might have (1)  $\mathfrak{A} = \mathfrak{N}$ , (2)  $\mathfrak{A} = (e_1, \mathfrak{N})$ , (3)  $\mathfrak{A} = (e_2, \mathfrak{N})$ , (4)  $\mathfrak{A} = (e_1 + e_2, \mathfrak{N})$ , or (5)  $\mathfrak{A} = (e_1, e_2, \mathfrak{N})$ .

EXAMPLE 2. Let  $\mathfrak{N}$  have a basis  $x_1, x_2, x_3, x_4$  where  $x_1^2 = x_2$ ,  $x_1x_2 = x_4$ ,  $x_1x_3 = 0$ ,  $x_1x_4 = 0$  and  $x_i^2 = x_ix_j = 0$  for  $i \neq 1$ .

Here

$$x_1 \rightarrow \left[ \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right],$$

$$x_2 \rightarrow \left[ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right],$$

while  $x_3$  and  $x_4$  are represented by 0 on both sides. Note that the radicals of  $[R(c)]$  and  $[L(c)]$  are not of the same order. By Theorem 3.1 the elements of  $[R(c)]$  are of the type

$$\begin{pmatrix} a_1 & b_1 & e_1 & f_1 \\ c_1 & d_1 & g_1 & h_1 \\ 0 & 0 & a_1 & b_1 \\ 0 & 0 & c_1 & d_1 \end{pmatrix},$$

and those of  $[L(c)]$  of the type

$$\begin{pmatrix} a_2 & b_2 & e_2 & f_2 \\ 0 & d_2 & g_2 & h_2 \\ 0 & 0 & a_2 & 0 \\ 0 & 0 & 0 & a_2 \end{pmatrix}.$$

Here aside from the matrix corresponding to  $x_1$ ,  $[L(c)]$  can contain the unit matrix  $I$ , an idempotent

$$E = \begin{pmatrix} 0 & b_2 & b_2 g_2 & b_2 h_2 \\ 0 & 1 & g_2 & h_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

the idempotent  $I - E$ , or only one of these. If we use the automorphism

$$\begin{aligned} x_1 &\rightleftharpoons x_1 - b_2 x_2 - b_2 g_2 x_3 - b_2 h_2 x_4, \\ x_2 &\rightleftharpoons x_2 + g_2 x_3 + h_2 x_4, \\ x_3 &\rightleftharpoons x_3 - b_2 x_4, \\ x_4 &\rightleftharpoons x_4, \end{aligned}$$

$E$  takes the simple form

$$E = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In case  $E$  (or  $I-E$ ) actually occurs in  $[L(c)]$ , then by Theorem 3.1  $[R(c)]$  is further restricted to matrices of the form

$$\begin{pmatrix} a_1 & 0 & e_1 & f_1 \\ 0 & d_1 & 0 & 0 \\ 0 & 0 & a_1 & 0 \\ 0 & 0 & 0 & d_1 \end{pmatrix}.$$

Here write

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

From here on it is fairly simple to enumerate the possible algebras bound to  $\mathfrak{R}$ .

*Case I.*  $\mathfrak{A}_d$  is void.

Here  $\mathfrak{A} = \mathfrak{A}_r + \mathfrak{A}_l + \mathfrak{R}$  and  $\mathfrak{A}_r$  and  $\mathfrak{A}_l$  are faithfully represented without remnants.

(a)  $\mathfrak{A}_l$  is void or  $I$ . Then  $\mathfrak{A}_r$  is void or any semisimple matrix algebra of the permissible form. These can be of order 1, 2, or 4.

(b)  $\mathfrak{A}_l$  contains  $E$ ,  $I-E$ , or both. Then  $\mathfrak{A}_r$  is void or contains  $F$ ,  $I-F$  or both.

*Case II.*  $\mathfrak{A}_d$  contains only one element independent of  $\mathfrak{R}$ . This element can be taken as an idempotent  $e$ .

(a)  $e$  is a left unit of  $\mathfrak{A}$ .

$\mathfrak{A}_l$  must be void. If  $e$  is a right unit, then  $\mathfrak{A}_r$  is void, and  $\mathfrak{A} = \mathfrak{A}_d = (1, \mathfrak{R})$ . We may also have, using automorphisms to simplify the form of the matrices,  $e \rightleftharpoons [F, I, 0]$ . Here  $\mathfrak{A}_d = (e, x_1, x_2)$  and  $\mathfrak{A}_r$  is void or has an idempotent  $f \rightleftharpoons [I-F, 0, 0]$ . It is also possible that  $e \rightleftharpoons [I-F, I, 0]$  and  $\mathfrak{A}_d = (e, x_2, x_4)$  while  $\mathfrak{A}_r$  is void or has an idempotent  $f \rightleftharpoons [F, 0, 0]$ .

(b)  $e$  is a right unit of  $\mathfrak{A}$ .

The possibilities here are similar to those above.

(c)  $e$  is neither a right nor a left unit of  $\mathfrak{A}$ .

Here  $e$  has  $E$  or  $I-E$  as its left representation and  $F$  or  $I-F$  as its right representation while  $\mathfrak{A}_r$  and  $\mathfrak{A}_l$  are void or contain the idempotent  $E$ ,  $I-E$ ,  $F$ ,  $I-F$  not representing  $e$ .

*Case III.*  $\mathfrak{A}_d$  contains two elements independent of  $\mathfrak{R}$ .

Here  $\mathfrak{A}_r$  and  $\mathfrak{A}_l$  are both void and  $\mathfrak{A} = \mathfrak{A}_d$ .  $\mathfrak{A}$  has a unit.  $\mathfrak{A} = (1, e, \mathfrak{R})$  where  $1 \rightleftharpoons [I, I, 0]$  and  $e \rightleftharpoons [F, E, 0]$  or  $[F, I-E, 0]$  or  $[I-F, E, 0]$  or  $[I-F, I-E, 0]$ .

Throughout this example, as a consequence of Theorem 3.5, all remnants may be taken as zero.



EXAMPLE 3.  $\mathfrak{R} = (x_1, x_2)$ ,  $\mathfrak{R}^2 = 0$ ,  $K$  of characteristic 2.

In a particular algebra bound to  $\mathfrak{R}$ ,

$$1 \rightleftharpoons [I, I, 0],$$

$$y \rightleftharpoons \left[ \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right],$$

where  $a$  is not a square in  $K$ . Here we may take  $r(1, 1) = r(1, y) = r(y, 1) = 0$ , but any value  $r(y, y) = bx_1 + cx_2$  will be permissible under equations (3.15), and since  $K$  is of characteristic 2 a change of representative will not affect the remnants. This is a case in which the center of one of the simple algebras of  $\mathfrak{A}/\mathfrak{R}$  is inseparable.

YALE UNIVERSITY,  
NEW HAVEN, CONN.

# INTEGERS OF QUADRATIC FIELDS AS SUMS OF SQUARES

BY  
IVAN NIVEN

**1. Introduction.** Lagrange proved that every positive rational integer is a sum of four squares of rational integers. Our principal result is that in an imaginary quadratic field every integer of the form

$$(1) \quad a + 2b\theta, \quad \theta^2 = -m,$$

$m$  being a positive square-free rational integer, is expressible as a sum of three squares of integers of the field. Gaussian integers are treated in §3, integers of the general imaginary quadratic field in §4; necessary and sufficient conditions for two-square sums are given in each case. Section 6 treats real quadratic integers, and §7 interprets some of the results in the theory of Diophantine equations.

It will be recalled that the coefficients of quadratic integers are not always rational integers. Specifically, if the field is an extension of the rational number field by  $\theta$  in equation (1), and if  $m \equiv 3 \pmod{4}$ , the integers of the field are given by

$$(2) \quad \frac{a}{2} + \frac{b}{2}\theta$$

where  $a$  and  $b$  are rational integers, both odd or both even. This introduces a special problem, which is dealt with for imaginary fields in §5. Roman letters represent rational integers throughout.

**2. Mordell's theorem.** In this section we prove a theorem which was stated by L. J. Mordell [1], and upon which most of our study is based. Mordell's proof contains an omission of such import that a complete proof is offered here.

**THEOREM 1.** *If  $f(x, y) = ax^2 + 2hxy + by^2$  is a positive binary quadratic form with integral coefficients, necessary and sufficient conditions that  $f$  be expressible as a sum of the squares of two linear forms with integral coefficients,*

$$(3) \quad f(x, y) = (a_1x + b_1y)^2 + (a_2x + b_2y)^2,$$

*are that  $\Delta = ab - h^2$  be a perfect square and that  $d = (a, h, b)$  have no prime factor of the form  $4n + 3$  to an odd power.*

To prove that these conditions are necessary, we take equation (3) as our

Presented to the Society, December 29, 1939; received by the editors December 12, 1939.

hypothesis and obtain

$$(4) \quad a = a_1^2 + a_2^2, \quad h = a_1b_1 + a_2b_2, \quad b = b_1^2 + b_2^2.$$

It follows that

$$\Delta = ab - h^2 = (a_1b_2 - a_2b_1)^2.$$

There is no loss of generality in assuming  $d$  to be square-free. Let  $p$  be a prime of the form  $4n+3$  dividing  $d$ , that is, dividing each of  $a$ ,  $b$ , and  $h$ . Using the theory of the decomposition of an integer into the sum of two squares, we note that the first and last equations of (4) imply that  $p$  is a divisor of  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$ . Hence  $p^2$  divides  $a$ ,  $b$ ,  $h$ , and therefore  $d$ , which contradicts our hypothesis that  $d$  is square-free.

Conversely, let us assume that

$$(5) \quad ab - h^2 = \Delta_0^2,$$

and that  $d$  is divisible by no prime of the form  $4n+3$  to an odd power, or, what is the same thing, that  $d$  is expressible as a sum of two squares of integers. Because of the identity

$$(U^2 + V^2)(u^2 + v^2) = (Uu + Vv)^2 + (Uv - Vu)^2,$$

and because we are attempting to prove that an equation of the form (3) can be set up, we may take

$$(6) \quad d = (a, h, b) = 1.$$

The gap in Mordell's argument occurs at this point. He states (page 5), "Now

$$ab - h^2 = \Delta_0^2, \quad h^2 \equiv -\Delta_0^2 \pmod{a},$$

and the solution of the congruence for  $h$  gives

$$h \equiv -\frac{\Delta_0 a_1}{a_2} \pmod{a}$$

for an appropriate resolution of  $a$  as a sum of two integral squares, say,

$$a = a_1^2 + a_2^2.$$

We shall show that  $a$  is expressible in the latter form, with

$$(7) \quad a_2 h \equiv -\Delta_0 a_1 \pmod{a}.$$

Let  $p$  be a prime of the form  $4n+3$  which divides  $ab$ . Equation (5) shows that  $ab$  is a sum of two squares; hence the highest power of  $p$  dividing  $ab$  is an even one, say  $p^{2\alpha}$ ; it follows that

$$p^a \mid h, \quad p^a \mid \Delta_0.$$

Equation (6) implies that  $p^{2a}$  divides  $a$  or  $b$ , but not both. Treating every prime factor of  $ab$  which is congruent to 3 modulo 4 in this fashion, we see that we may write

$$(8) \quad a = P^2 A, \quad b = Q^2 B, \quad h = PQH, \quad \Delta_0 = PQ\Delta_1,$$

wherein  $P$  and  $Q$  are odd, prime to each other, and contain only prime factors of the form  $4n+3$ ; also  $A$  and  $B$  contain no such prime factors. Equation (5) shows that

$$(9) \quad AB = H^2 + \Delta_1^2 = (H + \Delta_1 i)(H - \Delta_1 i).$$

Now each prime factor of  $A$  is expressible as a sum of two squares in one and only one way, so that we have

$$(10) \quad A = \prod_{j=1}^n (x_j^2 + y_j^2) = \prod_{j=1}^n (x_j + y_j i)(x_j - y_j i).$$

Each of the complex factors in the latter product is a prime in the field  $R(i)$ , the rational number field extended by  $i$ . The unique factorization law holds in  $R(i)$  so that  $x_j + y_j i$  divides one of the two factors

$$H + \Delta_1 i, \quad H - \Delta_1 i$$

of  $AB$ , and  $x_j - y_j i$  divides the other. Combining the terms of the product (10) according to this distinction, we may write

$$(11) \quad A = (A_1 - A_2 i)(A_1 + A_2 i) = A_1^2 + A_2^2,$$

where

$$(A_1 - A_2 i) \mid (H + \Delta_1 i), \quad (A_1 + A_2 i) \mid (H - \Delta_1 i).$$

Similarly we have

$$(12) \quad B = B_1^2 + B_2^2 = (B_1 - B_2 i)(B_1 + B_2 i),$$

these factors dividing  $H + \Delta_1 i$  and  $H - \Delta_1 i$  respectively. Equations (9), (11), and (12) imply

$$H + \Delta_1 i = (A_1 - A_2 i)(B_1 - B_2 i),$$

whence we obtain

$$(13) \quad H = A_1 B_1 - A_2 B_2, \quad \Delta_1 = -A_1 B_2 - A_2 B_1.$$

If we write

$$a_1 = PA_1, \quad a_2 = PA_2, \quad b_1 = QB_1, \quad b_2 = QB_2,$$

then equations (8), (11), (12), and (13) imply

$$(14) \quad a = a_1^2 + a_2^2, \quad b = b_1^2 + b_2^2, \quad h = a_1b_1 - a_2b_2, \quad \Delta_0 = -a_1b_2 - a_2b_1.$$

It follows that

$$a_2h = a_2a_1b_1 - a_2^2b_2,$$

$$a_2h \equiv a_2a_1b_1 + a_1^2b_2 \pmod{a},$$

and

$$a_2h \equiv -a_1\Delta_0 \pmod{a},$$

which we set out to prove. In fact, equations (14) imply

$$(15) \quad \frac{a_2h + a_1\Delta_0}{a} = -b_2, \quad \frac{a_1h - a_2\Delta_0}{a} = b_1,$$

and it is easily verified that

$$ax^2 + 2hxy + by^2 = (a_1x + b_1y)^2 + (a_2x - b_2y)^2.$$

**3. Gaussian integers.** Let us consider  $a + 2bi$ , where  $a$  and  $b$  are rational integers. We have, for an arbitrary integer  $t$ ,

$$a + 2bi = (a + t) + 2bi + ti^2$$

which may thus be considered as a quadratic form in 1 and  $i$ . Mordell's theorem is applicable; first we wish to show that there exists a rational integral value of  $t$  such that  $t(a+t) - b^2$  is a perfect square.

First let  $a$  be even,  $a = 2A$ . We wish to obtain integral  $t$  and  $x$  to satisfy

$$t(2A + t) - b^2 = x^2,$$

which may be written in the form

$$(16) \quad (t + A)^2 - x^2 = A^2 + b^2.$$

This equation has no solutions if both  $A$  and  $b$  are odd, but is solvable otherwise.

In case either  $A$  or  $b$  is odd, we write the solution

$$t + A + x = A^2 + b^2, \quad t + A - x = 1,$$

so that

$$t = \frac{(A - 1)^2 + b^2}{2}.$$

In our application of Theorem 1, we have (in the case considered) satisfied the condition that the negative of the discriminant of the form be a square.



We now consider the nature of  $d$ , the greatest common divisor of  $t$ ,  $b$ , and  $a+t$ . The above equations show that

$$d = \left( b, \frac{(A-1)^2 + b^2}{2}, \frac{(A+1)^2 + b^2}{2} \right).$$

Let  $p$  be any odd prime dividing  $d$ . It is an immediate consequence of the above equation that  $p$  divides  $b$ ,  $A-1$ , and  $A+1$ . Hence  $p=1$ , and  $d$  is not divisible by any odd prime.

In case both  $A$  and  $b$  are even, we write  $A=2A_1$ ,  $b=2b_1$ , and equation (16) has the solutions

$$t + 2A_1 + x = 2(A_1^2 + b_1^2), \quad t + 2A_1 - x = 2,$$

so that

$$t = (A_1 - 1)^2 + b_1^2.$$

The value of  $d$  is now given by the equation

$$d = (b, (A_1 - 1)^2 + b_1^2, (A_1 + 1)^2 + b_1^2).$$

The argument of the last paragraph applies again to show that  $d$  is divisible by no odd prime. This completes the discussion when  $a$  is even.

In case  $a$  is odd,  $a=2A+1$ , equation (16) is replaced by

$$(17) \quad (2t + a)^2 - 4x^2 = a^2 + 4b^2.$$

This equation always has rational integral solutions  $t$  and  $x$ . For if we write

$$2t + a + 2x = a^2 + 4b^2, \quad 2t + a - 2x = 1,$$

the solution is

$$t = A^2 + b^2.$$

Again we see that  $d$  is divisible by no odd prime, because

$$d = (b, A^2 + b^2, (A+1)^2 + b^2).$$

Recalling the remark after equation (16), we have shown that  $a+2bi$  is expressible as a quadratic form in 1 and  $i$  satisfying the conditions of Theorem 1 provided that not both  $a/2$  and  $b$  are integral and odd. Hence if these conditions on  $a$  and  $b$  are satisfied, the integer  $a+2bi$  is expressible as a sum of two squares of Gaussian integers.

Conversely, suppose that the Gaussian integer  $a+2bi$  is a sum of two squares,

$$a + 2bi = (c + di)^2 + (e + fi)^2.$$

Setting  $t = d^2 + f^2$ , we have the result

$$(a + t)x^2 + 2bxy + ty^2 = (cx + dy)^2 + (ex + fy)^2.$$

Theorem 1 shows that  $t(a+t) - b^2$  must be the square of an integer, and the conditions for equations (16) or (17) must be fulfilled for  $a$  even or odd, respectively. But equation (16) cannot be satisfied if  $a/2$  and  $b$  are odd integers. Hence the Gaussian integer  $a + 2bi$  is not expressible as a sum of two squares if  $a/2$  and  $b$  are odd rational integers. We have proven the first statement of the following theorem.

**THEOREM 2.** *A Gaussian integer of the form  $a + 2bi$  is expressible as a sum of two squares of Gaussian integers if and only if not both  $a/2$  and  $b$  are odd integers. Every Gaussian integer of the form  $a + 2bi$  is expressible as a sum of three squares. A Gaussian integer is expressible as a sum of squares of Gaussian integers if and only if its imaginary coordinate is even.*

The last remark is trivial. The second statement is a corollary of the first. For if  $a/2$  and  $b$  are integral and odd, the integer  $a - 1 + 2bi$  is expressible as a sum of two squares, whence  $a + 2bi$  is a sum of three squares, one of which is unity.

**4. General imaginary quadratic fields.** We now consider integers of the form  $\gamma = a + 2b\theta$  where  $a$  and  $b$  are rational integers, and

$$(18) \quad \theta^2 = -m,$$

$m$  being an integer greater than unity with no square factors. We note that  $\gamma$  is expressible in infinitely many ways as a quadratic form in 1 and  $\theta$ ,

$$\gamma = (a + tm) + 2b\theta + t\theta^2,$$

$t$  being an arbitrary integer. If  $t$  can be selected so that this quadratic form is expressible as a sum of two squares of linear forms, then  $\gamma$  is a sum of two squares of integers of the field  $R(\theta)$ .

On the other hand, if  $a + 2b\theta$  is a sum of two squares,

$$a + 2b\theta = (c + d\theta)^2 + (e + f\theta)^2,$$

we set  $t = d^2 + f^2$  as before and obtain

$$(a + tm)x^2 + 2bxy + ty^2 = (cx + dy)^2 + (ex + fy)^2.$$

We have shown, therefore, that the integer  $\gamma$  is a sum of two squares of integers of  $R(\theta)$  with rational integral coordinates if and only if the quadratic form  $[a + tm, 2b, t]$  is expressible as a sum of two squares of linear forms by means of a suitable choice of the rational integer  $t$ . Hence Theorem 1 is applicable.

**THEOREM 3.** *The integer  $a + 2b\theta$  is expressible as the sum of the squares of*

two integers of the form  $c+d\theta$ , if and only if there exists an integer  $t$  such that

$$mt^2 + at - b^2$$

is a perfect square, and such that  $(t, b, a+mt)$  is not divisible by a prime of the form  $4n+3$  to an odd power.

We now consider the problem of expressing the integer  $a+2b\theta$  as a sum of three squares. The equation

$$(19) \quad a + 2b\theta - (u + v\theta)^2 = (tm + a - u^2) + 2(b - uv)\theta + (t - v^2)\theta^2$$

leads us to search for integral values of  $t$ ,  $u$ , and  $v$  which will make

$$(20) \quad (t - v^2)(tm + a - u^2) - (b - uv)^2$$

a perfect square. First we set the terms free from  $t$  equal to a square,

$$v^2(u^2 - a) - (b - uv)^2 = y^2,$$

so that

$$(21) \quad u = \frac{v^2a + b^2 + y^2}{2bv}.$$

To obtain an integer from this expression for  $u$ , we set  $v=b$  and  $y=2Yb$  or  $y=(2Y+1)b$  according as  $a$  is odd or even; the integer  $Y$  is arbitrary.

The expression (20) is written

$$mt^2 + t(a - u^2 - mb^2) + y^2 = \left(y - \frac{pt}{q}\right)^2,$$

and a solution is

$$(22) \quad t = q(2py + aq - u^2q - mb^2q),$$

provided

$$(23) \quad p^2 - mq^2 = 1.$$

Note that  $pt/q$  is an integer.

We shall also need to account for the greatest common divisor of the coefficients of the quadratic form on the right side of equation (19),

$$(24) \quad (b - uv, a - u^2 + tm, t - v^2) = (b - ub, a - u^2 + tm, t - b^2).$$

LEMMA. Let  $\pi_1, \pi_2, \dots, \pi_r$  be the primes of the form  $4n+3$  which divide  $b$ . Then we can choose  $y$  in (21) so that

$$(25) \quad u^2 \not\equiv a \pmod{\pi_j}, \quad j = 1, 2, \dots, r.$$

First consider  $a$  odd,  $a=2A-1$ . Then  $y=2Yb$ , and (21) becomes

$$u = A + 2Y^2.$$

In this case (25) becomes

$$(26) \quad (A + 2Y^2)^2 \not\equiv 2A - 1 \pmod{\pi_j}, \quad j = 1, 2, \dots, r.$$

When  $Y$  ranges over a complete residue system modulo  $\pi_j$ ,  $Y^2$  (and therefore  $2Y^2 + A$ ) takes on  $\frac{1}{2}(\pi_j + 1)$  incongruent values modulo  $\pi_j$ . From the theory of quadratic residues it follows that  $(2Y^2 + A)^2$  takes on at least  $[\frac{1}{2}(\pi_j + 1)]$  incongruent values modulo  $\pi_j$ , where  $[x]$  has the usual number-theoretic meaning, namely, the greatest integer less than or equal to  $x$ . Since  $[\frac{1}{2}(\pi_j + 1)] \geq 2$  for all primes greater than 5, it is possible to select a value  $s_j$  from the complete system of residues modulo  $\pi_j$ , so that (26) is satisfied for all primes greater than 5 provided

$$(27) \quad Y \equiv s_j \pmod{\pi_j} \quad (j = 1, \dots, r), \pi_j > 5.$$

Since 5 is not a prime of the form  $4n+3$ , we take  $\pi_j = 3$  as the special case. In this case, choose  $Y \equiv 0, 1, 2 \pmod{3}$  when  $A \equiv 0, 1, 2 \pmod{3}$  respectively, and (25) is satisfied.

Thus an  $s_j$  can be found corresponding to each  $\pi_j$  in (27) including the case  $\pi_j = 3$  if it happens to be present, so that values of  $Y$  satisfying (26) may be found by use of the Chinese remainder theorem. Hence the lemma is proven in case  $a$  is odd.

If  $a = 2A$ , we have  $y = (2Y+1)b$ ; equations (21) and (25) become

$$u = A + 1 + 2Y^2 + 2Y,$$

and

$$(28) \quad (A + 1 + 2Y^2 + 2Y)^2 \not\equiv 2A \pmod{\pi_j}, \quad j = 1, \dots, r,$$

respectively. Again let  $Y$  range over a complete residue system modulo  $\pi_j$ ; each of the quantities  $Y^2 + Y$  and  $2Y^2 + 2Y + A + 1$  takes on  $\frac{1}{2}(\pi_j + 1)$  incongruent values modulo  $\pi_j$ . Thus the expression

$$(2Y^2 + 2Y + A + 1)^2$$

takes on at least  $[\frac{1}{2}(\pi_j + 1)]$  incongruent values modulo  $\pi_j$ . As in the earlier case, a relation of the type (27) is established. In case the prime under discussion is 3, equation (28) is satisfied by choosing  $Y \equiv 0, 1, 2 \pmod{3}$  when  $A \equiv 0, 1, 2 \pmod{3}$  respectively. The proof of the lemma is completed by use of the Chinese remainder theorem, as in the previous case.

Having thus chosen a suitable value of  $u$ , we note that  $q$  in (23) may be selected so that it is divisible by  $b(u-1)$ . For we may set

$$(29) \quad q = b(u-1)Q$$

where  $Q$  is a solution of

$$(30) \quad p^2 - mb^2(u-1)^2Q^2 = 1.$$

This is a Pell equation, and is known to have solutions  $p$  and  $Q$  because  $mb^2(u-1)^2$  is not a square.

It is not difficult to show that the expression on the right side of equation (24) has no prime of the form  $4n+3$  as a factor. For suppose that  $\pi$  is such a prime dividing  $b-ub$ . Equation (29) shows that  $\pi$  divides  $q$ , and consequently equation (22) implies that  $\pi$  divides  $t$ . First, if  $\pi$  divides  $b$ , the lemma states that  $\pi$  does not divide  $u^2-a$ , and hence  $a-u^2+tm$  is prime to  $\pi$ . On the other hand, if  $\pi$  divides  $u-1$  but not  $b$ , it is clear that  $\pi$  cannot divide the expression  $t-b^2$  in equation (24).

We have satisfied the conditions of Theorem 1, and equation (19) may therefore be interpreted as follows:

**THEOREM 4.** *Every integer of the form  $a+2b\theta$  of the quadratic field  $R(\theta)$  defined by equation (18) is expressible as a sum of three squares of integers of the field.*

**5. A special case.** We now examine more thoroughly the fields  $R(\theta)$  where the integer  $m$  of equation (19) is of the form  $4n+3$ .

**THEOREM 5.** *In case  $m \equiv 3 \pmod{4}$ , the integer  $a+b\theta$ , with  $b$  an odd rational integer, is expressible as a sum of two squares of integers of the field if and only if  $4a+4b\theta$  is expressible as a sum of two squares of integers of the type  $c+d\theta$  (see Theorem 3). Also, the integer  $\frac{1}{2}a+\frac{1}{2}b\theta$ , with  $a$  and  $b$  odd rational integers, is expressible as a sum of two squares of integers of the field if and only if  $2a+2b\theta$  is expressible as a sum of two squares of integers of the type  $c+d\theta$ .*

It is obvious that each of these conditions is necessary for the proposed representation. To show that the condition expressed in the first statement is sufficient, we assume that  $4a+4b\theta$  is a sum of two squares,

$$(31) \quad 4a + 4b\theta = (x_1 + y_1\theta)^2 + (x_2 + y_2\theta)^2.$$

This implies the congruence

$$0 \equiv x_1^2 + x_2^2 - my_1^2 - my_2^2 \pmod{4},$$

or

$$0 \equiv x_1^2 + x_2^2 + y_1^2 + y_2^2 \pmod{4}.$$

Hence every one of  $x_1, x_2, y_1$  and  $y_2$  is even, or every one is odd. Equation (31) can be divided by 4 to give the desired result.

We now turn to the second statement of the theorem and assume that

$$(32) \quad 2a + 2b\theta = (x_1 + y_1\theta)^2 + (x_2 + y_2\theta)^2.$$

Since  $a$  and  $b$  are odd, we obtain the congruences

$$2 \equiv x_1^2 + x_2^2 + y_1^2 + y_2^2 \pmod{4}, \quad 1 \equiv x_1y_1 + x_2y_2 \pmod{2}.$$



These imply that  $x_1$  and  $y_1$  are both even or both odd, and an analogous result for  $x_2$  and  $y_2$ . The equation (32) can be divided by 4 to give the desired result, and we have proven the theorem.

**THEOREM 6.** *In case  $m \equiv 3 \pmod{4}$ , every integer of the field  $R(\theta)$  is expressible as a sum of three squares of integers of the field.*

Because of Theorem 4 we need consider only integers of the types

$$(33) \quad a + b\theta, \quad b \equiv 1 \pmod{2},$$

and

$$(34) \quad \frac{a}{2} + \frac{b}{2}\theta, \quad a \equiv b \equiv 1 \pmod{2}.$$

By Theorem 4 we have

$$(35) \quad 4a + 4b\theta = \sum_{j=1}^3 (x_j + y_j\theta)^2,$$

from which we obtain the congruences

$$\begin{aligned} 0 &\equiv x_1^2 + x_2^2 + x_3^2 + y_1^2 + y_2^2 + y_3^2 \pmod{4}, \\ 0 &\equiv x_1y_1 + x_2y_2 + x_3y_3 \pmod{2}. \end{aligned}$$

If there were a disparity between  $x_1$  and  $y_1$  with respect to 2, these congruences would imply

$$\begin{aligned} 3 &\equiv x_2^2 + x_3^2 + y_2^2 + y_3^2 \pmod{4}, \\ 0 &\equiv x_2y_2 + x_3y_3 \pmod{2}, \end{aligned}$$

which have no solutions in integers. Hence  $x_1$  and  $y_1$  are both odd or both even, and an analogous argument holds for the pairs  $x_2, y_2$  and  $x_3, y_3$ . Our theorem is proven for integers of types (33) by dividing equation (35) by 4.

Turning to integers of the type (34), we write the equation

$$(36) \quad 2a + 2b\theta = \sum_{j=1}^3 (x_j + y_j\theta)^2,$$

using Theorem 4 as our authority. This equation implies the congruences

$$\begin{aligned} 2 &\equiv x_1^2 + x_2^2 + x_3^2 + y_1^2 + y_2^2 + y_3^2 \pmod{4}, \\ 1 &\equiv x_1y_1 + x_2y_2 + x_3y_3 \pmod{2}. \end{aligned}$$

If  $x_1$  and  $y_1$  are incongruent modulo 2, we would have

$$1 \equiv x_2^2 + x_3^2 + y_2^2 + y_3^2 \pmod{4}, \quad 1 \equiv x_2y_2 + x_3y_3 \pmod{2},$$

which are manifestly impossible in integers. Hence  $x_1$  and  $y_1$  are both even or both odd; a similar statement holds for the pairs  $x_2, y_2$  and  $x_3, y_3$ . Dividing equation (36) by 4, we have completed the proof of the theorem.

6. Real quadratic fields. Let

$$f(x, y) = ax^2 + 2hxy + by^2$$

be a positive form with integral coefficients. Mordell [2] has shown that  $f$  is expressible as a sum of five squares of linear forms with integral coefficients; also he has shown that  $f$  is expressible as a sum of four squares of linear forms with integral coefficients if and only if  $ab - h^2$  is a sum of three squares of integers, that is, if and only if  $ab - h^2$  is not of the form  $4^r(8s+7)$ . In case the expression  $ab - h^2$  equals zero, the form  $f$  is expressible as a sum of four squares of linear forms with integral coefficients.

Let us now consider the field  $R(m^{1/2})$ , where  $m$  is a square-free rational integer greater than unity. The integer  $a + 2bm^{1/2}$  can be written in the form

$$(37) \quad a + 2bm^{1/2} = (a - tm) + 2bm^{1/2} + t(m^{1/2})^2,$$

a quadratic form in 1 and  $m^{1/2}$ . If Mordell's theorems above are to apply, we must first inquire whether  $t$  can be chosen so that the right side of equation (37) is a *positive* form. The question is whether a positive value of  $t$  can be chosen so that

$$(38) \quad D = (a - tm)t - b^2 > 0.$$

If we define  $K$  by the equation

$$(39) \quad K = (a^2 - 4mb^2)^{1/2},$$

it is seen that  $D$  vanishes when  $t$  has the values

$$\frac{a - K}{2m}, \quad \frac{a + K}{2m}.$$

Furthermore, if  $K$  is real, and if  $t$  lies between these values, then  $t$  and  $D$  are positive, and the right side of equation (37) is a positive form; hence the first Mordell theorem stated at the beginning of this section is applicable. On the other hand, if  $t$  equals one of the above values, being real, then  $D$  is zero, and we apply the last Mordell theorem stated.

**THEOREM 7.** *The integer  $a + 2bm^{1/2}$  is expressible as a sum of five squares of integers of the form  $c + dm^{1/2}$  if and only if the quantity  $K$  defined by (39) is real and the closed interval*

$$(40) \quad \left( \frac{a - K}{2m}, \frac{a + K}{2m} \right)$$

*contains a rational integer.*

Since any integer contained in the interval (40) is of necessity positive, it is clear that these conditions are sufficient. Conversely, let us assume that there exist integral values  $x_j, y_j$  ( $j=1, \dots, 5$ ) such that

$$a + 2bm^{1/2} = \sum_{j=1}^5 (x_j + y_j m^{1/2})^2,$$

from which we obtain

$$a = \sum_{j=1}^5 (x_j^2 + m y_j^2), \quad b = \sum_{j=1}^5 x_j y_j.$$

The equation

$$a - 2bm^{1/2} = \sum_{j=1}^5 (x_j - y_j m^{1/2})^2$$

shows that  $a^2 - 4mb^2$  is positive, and consequently  $K$  in (39) is real. Consider the function

$$D = -mt^2 + at - b^2,$$

$t$  being looked upon as a continuous variable. Its graph is a parabola. Its zeros are the end-points of the interval (40). Furthermore, any value of  $t$  for which  $D$  is positive lies in the interval (40). When  $t$  is given the integral value  $\sum_{j=1}^5 y_j^2$ ; we obtain

$$D = \sum_{j=1}^5 x_j^2 \sum_{j=1}^5 y_j^2 - \left( \sum_{j=1}^5 x_j y_j \right)^2.$$

By the elementary theory of inequalities, this is not negative. Hence we have exhibited an integral value satisfying the conditions of the theorem.

**THEOREM 8.** *The integer  $a + 2bm^{1/2}$  is expressible as a sum of four squares of integers of the form  $c + dm^{1/2}$  if and only if  $K$  defined by (39) is real and the closed interval (40) contains a rational integer  $t$  so that the value of  $D$  in equation (38) is expressible as a sum of three squares of rational integers.*

This theorem needs no explanation, since it is an immediate extension of Theorem 7, obtained by the use of Mordell's work as outlined at the beginning of this section. It is also possible to state theorems analogous to the last theorem for the situations wherein we wish two-square and three-square sums; this would be done by use of Theorem 1 and other work [3] of Mordell.

**7. Consequences in the theory of Diophantine equations.** The first statement of Theorem 2 may be interpreted as follows:

**THEOREM 9.** *The Diophantine equations*

$$x^2 + y^2 - z^2 - w^2 = a, \quad xz + yw = b,$$

are solvable simultaneously if and only if not both  $\frac{1}{2}a$  and  $b$  are integral and odd.

The second statement of Theorem 2 together with Theorem 4 leads to the following result.

**THEOREM 10.** *If  $a$  and  $b$  are arbitrary integers, and if  $m$  is unity or an integer greater than unity which is not a square, then the equations*

$$\begin{aligned}x^2 + y^2 + z^2 - m(w^2 + u^2 + v^2) &= a, \\ xw + yu + zv &= b,\end{aligned}$$

are solvable simultaneously in integers.

Theorem 4 was proven with  $m$  a square-free integer, but the proof is valid with the less restrictive hypothesis that  $m$  be no square. This hypothesis is needed to insure solutions for the Pell equation (30). Finally we rewrite Theorem 7.

**THEOREM 11.** *If  $a$  and  $b$  are arbitrary integers, and if  $m$  is any positive integer, then the equations*

$$\sum_{j=1}^b (x_j^2 + my_j^2) = a, \quad \sum_{j=1}^b x_j y_j = b$$

have simultaneous solutions in integers if and only if the quantity  $K$  defined by equation (39) is real and the closed interval (40) contains a rational integer.

The restriction that  $m$  be square-free contained in Theorem 7 is abandoned here because it was not used in the proof. It was included in Theorem 7 merely because quadratic integers are defined in terms of a square-free rational integer.

#### REFERENCES

1. L. J. Mordell, *On the representation of a binary quadratic form as a sum of squares of linear forms*, Mathematische Zeitschrift, vol. 35 (1932), pp. 1-15.
2. ———, *A new Waring's problem*, Quarterly Journal of Mathematics, vol. 1 (1930), pp. 276-288.
3. ———, *On binary quadratic forms*, Journal für die reine und angewandte Mathematik, vol. 167 (1932).

UNIVERSITY OF ILLINOIS,  
URBANA, ILL.

## ON FINITELY MEAN VALENT FUNCTIONS. II

BY

D. C. SPENCER

1. We suppose  $f(z)$  is regular in  $|z| < 1$  and denote by  $W$  the Riemann domain which is the transform of  $|z| < 1$  by  $f$ . We shall say that  $f(z)$  has valency  $p$  if  $f(z)$  takes no value  $w$  more than  $p$  times. More generally, let  $W(R)$  be the area (regions covered multiply being counted multiply) of that portion of  $W$  which lies in the circle  $|w| \leq R$ ; then, if

$$(1.1) \quad W(R) \leq p\pi R^2$$

for all  $R > 0$ , where  $p$  is a positive number (not necessarily integral), we shall say that  $f(z)$  is  $p$  mean valent (p.m.v.)<sup>(1)</sup>. This paper is a sequel to one of the same title to appear shortly in the Proceedings of the London Mathematical Society<sup>(2)</sup> in which I have shown that many of the known theorems concerning  $p$ -valent functions may be extended to the wider class of p.m.v. functions. I discuss here the behavior of p.m.v. functions on paths tending to points on the circumference  $|z| = 1$ .

The theorems which I discuss here remain true under hypotheses somewhat less restrictive than the one stated above. For example, the hypothesis that  $W(R) \leq p\pi R^2$  only for  $R \geq R_0 > 0$  would suffice (constants now depending on  $R_0$  as well as  $p$ ). Furthermore, slightly less precise versions of the theorems (with  $p$  replaced by  $p + \epsilon$ ) could be stated subject to the still weaker condition that

$$\limsup_{R \rightarrow \infty} \frac{W(R)}{\pi R^2} \leq p.$$

Certain theorems<sup>(3)</sup> proved elsewhere, however, require the full strength of (1.1) for all  $R > 0$ , and for this reason I have not introduced a new definition here.

2. We begin by expressing the inequality (1.1) in a form more convenient for our purpose. Let  $n(r, w)$  be the number of times (necessarily bounded by a constant depending on  $r$ ) that  $f(z)$  takes on the value  $w$  in  $|z| < r$ ; and let us take

---

Presented to the Society, April 27, 1940; received by the editors October 13, 1939, and, in expanded form, April 4, 1940.

(<sup>1</sup>) This definition was suggested to me by Professor J. E. Littlewood, to whom I am also indebted for advice in the preparation of the paper.

(<sup>2</sup>) This paper will be referred to as  $V_1$ .

(<sup>3</sup>) For example, Theorem 1 of  $V_1$ . The complication of an additional parameter  $R_0$  is avoided thereby as well.



$$(2.1) \quad p(r, R) = \frac{1}{2\pi} \int_{-\pi}^{\pi} n(r, Re^{i\psi}) d\psi,$$

$$(2.2) \quad p(R) = p(1, R) = \lim_{r \rightarrow 1} p(r, R).$$

Since  $p(r, R)$  is an increasing function of  $r$ ,  $p(R)$  exists (but may be infinite). We have

$$\begin{aligned} W(R) &= \lim_{r \rightarrow 1} \int_0^R \int_{-\pi}^{\pi} n(r, Re^{i\psi}) R dR d\psi \\ &= \int_0^R \left( \lim_{r \rightarrow 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} n(r, Re^{i\psi}) d\psi \right) d(\pi R^2) \\ &= \int_0^R p(R) d(\pi R^2). \end{aligned}$$

Hence the hypothesis (1.1) may be expressed in the form

$$(2.3) \quad \int_0^R p(R) d(\pi R^2) \leq p\pi R^2 \quad (R > 0).$$

3. We shall make frequent use of the following lemma:

LEMMA 1. Suppose  $s_1 \geq s_2$ . Then the hypothesis

$$(3.1) \quad \int_0^{R_1} p(R) d(R^{s_1}) \leq pR_1^{s_1} \quad (R_1 > 0)$$

implies

$$(3.2) \quad \int_0^{R_1} p(R) d(R^{s_2}) \leq pR_1^{s_2} \quad (R_1 > 0),$$

but not conversely.

Making some trivial transformations of variable, we see it is enough to show that, if  $s \geq 1$ ,

$$(3.3) \quad \int_0^{R_1} p_1(R) d(R^s) \leq pR_1^s \quad (R_1 > 0)$$

implies

$$(3.4) \quad \int_0^{R_1} p_1(R) dR \leq pR_1 \quad (R_1 > 0),$$

where  $p_1(R) = p(R^{1/s})$ , but that (3.4) does not imply (3.3).

Integrating by parts, we have (dropping subscripts)

$$\begin{aligned}\int_0^{R_1} p(R) dR &= \int_0^{R_1} p(R) \cdot R^{s-1} \cdot R^{1-s} dR \\ &= \left[ R^{1-s} \int_0^R p(R) \cdot R^{s-1} dR \right]_0^{R_1} \\ &\quad + (s-1) \int_0^{R_1} \left( \int_0^R p(R) R^{s-1} dR \right) \cdot R^{-s} dR \\ &\leq \frac{1}{s} p R_1 + \frac{(s-1)}{s} p R_1 \\ &= p(R_1)\end{aligned}$$

by (3.3).

On the other hand, the converse implication is false. In fact, take

$$(3.5) \quad p(R) = \begin{cases} 1, & 2\mu - 1 \leq R < 2\mu, \mu = 1, 2, \dots, \\ 0, & \text{otherwise;} \end{cases}$$

and write  $R_1 = n + \theta$ , where  $n$  is an integer and  $0 \leq \theta < 1$ . Then

$$\begin{aligned}\int_0^{R_1} p(R) dR &= \sum_{\mu=1}^{[n/2]} \{ (2\mu - 1) - 2\mu \} + \begin{cases} 0, & n \text{ even,} \\ \theta, & n \text{ odd,} \end{cases} \\ &= \begin{cases} \frac{1}{2}n, & n \text{ even,} \\ \frac{1}{2}(n-1) + \theta, & n \text{ odd,} \end{cases} \\ &\leq \frac{1}{2}R_1.\end{aligned}$$

Hence (3.4) is satisfied with  $p = \frac{1}{2}$ . But

$$\begin{aligned}\int_0^{2\nu} p(R) d(R^s) &= \int_0^{(2\nu)^s} p(R^{1/s}) dR = \sum_{\mu=1}^{\nu} \{ (2\mu)^s - (2\mu-1)^s \} \\ &= \frac{1}{2}(2\nu)^s + \frac{1}{4}s(2\nu)^{s-1} + O(\nu^{s-2}) > \frac{1}{2}(2\nu)^s,\end{aligned}$$

if  $s > 1$  and  $\nu > \nu_0(s)$ . Thus, if  $s > 1$ , (3.3) is false for  $R_1 = 2\nu$ ,  $\nu > \nu_0$ . We have shown that the converse of the lemma is false for some function  $p(R)$ , but not for a  $p(R)$  corresponding to an actual Riemann domain. However, the  $p(R)$  of the schlicht function which maps the unit circle on the domain shown in Fig. 1 differs as little as we please from the choice (3.5), and for it, therefore, the converse of the lemma is false.

4. Lemma 1 shows that the hypothesis

$$(s) \quad \int_0^{R_1} p(R) d(R^s) \leq p R_1^s \quad (R_1 > 0)$$

is the stronger the larger  $s$  is. For the sake of completeness I include the following two theorems (but they may be omitted by the reader if he so desires; they have no bearing on the rest of the theory).

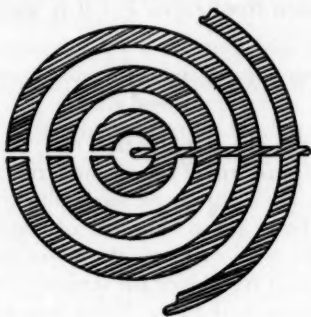


FIG. 1

**THEOREM 1.** *If (s) is true for all  $s > 0$ , and  $p(R)$  corresponds to a Riemann domain  $W^{(4)}$ , then  $p(R) \leq p$ .*

**THEOREM 2.** *If*

$$f(z) = a_1 z + a_2 z^2 + \dots$$

*is mean  $p$ -valent, so is the (generally algebraic<sup>(5)</sup>) function  $\{f(z^k)\}^{1/k}$ . On the other hand, if  $k > 1$ , the mean  $p$ -valency of the function*

$$f_k(z) = a_1 z + a_{k+1} z^{k+1} + a_{2k+1} z^{2k+1} + \dots$$

*does not imply that of the function (of form  $f_1$ )*

$$\{f_k(z^{1/k})\}^k = a_1^k z + k a_1^{k-1} a_{k+1} z^2 + \dots$$

If  $f_k(z)$  is  $p$ -valent, then so is  $\{f_k(z^{1/k})\}^k$ . This result and its converse are well known when the functions are  $p$ -valent<sup>(6)</sup>.

We take Theorem 1 first, and note that if for a given value of  $R$ ,  $R_0$  say,

$$p(R_0) = p(1, R_0) > p,$$

then, since  $p(r, R_0) \rightarrow p(R_0)$  as  $r \rightarrow 1$ , there exists a  $\delta > 0$  and  $r_0 = r_0(\delta) < 1$  such that, for  $r > r_0$ ,

<sup>(4)</sup> The theorem is false if this clause is omitted (and is therefore not trivial).

<sup>(5)</sup>  $\{f(z^k)\}^{1/k}$  has branch points at the zeros of  $f$  other than the origin. In the neighborhood of the origin, however,

$$\{f(z^k)\}^{1/k} = a_1^{1/k} z + (1/k) a_1^{1/k-1} a_{k+1} z^{k+1} + \dots$$

If  $f$  is mean 1-valent, then  $f$  has at most one zero (by the definition of mean 1-valency), and in this case, therefore,  $\{f(z^k)\}^{1/k}$  is regular in  $|z| < 1$  (and so of the form  $f_k$ ).

<sup>(6)</sup> See V<sub>1</sub>.

$$(4.1) \quad p(r, R_0) > p + \delta.$$

We show that this cannot happen.

Suppose it does. Then in the first place  $p(r, R)$  is discontinuous at  $R = R_0$ , *qua* function of  $R$ , for each fixed  $r > r_0$ . For if it were continuous we should have<sup>(7)</sup>

$$\lim_{s \rightarrow R_0} \frac{1}{R_0} \int_0^{R_0} p(r, R) d(R^s) = \lim_{s \rightarrow R_0} \int_0^1 p(r, xR_0) d(x^s) = p(r, R_0),$$

which is incompatible with the combination (4.1) and (s) for  $R = R_0$ .

Now let  $B(r)$  be the transform of the circumference  $|z| = r$  by  $f(z)$  ( $B(r)$  is the boundary of  $W(r)$ ). Then  $B(r)$  is an analytic curve, and crosses the circumference  $|w| = R_0$  a finite (even) number of times if it crosses it at all. If  $B(r)$  does not meet  $|w| = R_0$ , or meets it only in points, then it is obvious that  $p(r, R)$  is continuous at  $R_0$ . Hence the intersection of  $B(r)$  with  $|w| = R_0$  contains one or more intervals if  $r > r_0$ . These intervals depend upon  $r$ , but by (4.1) the intervals corresponding to any  $r > r_0$  have positive total length. It follows that if  $r > r_0$  the plane measure of  $B(r)$  is positive, and so the length (or linear measure) of  $B(r)$  is infinite. This is a contradiction of the regularity of  $f(z)$  in  $|z| < 1$ , and proves Theorem 1.

Next, to prove the first part of Theorem 2, let  $p(R)$ ,  $p_k(R)$  correspond respectively to  $f(z)$ ,  $\{f(z^k)\}^{1/k}$ . Then

$$p_k(R) = p(R^k).$$

In fact,  $\{f(z)\}^{1/k}$  (or branch thereof) maps  $|z| < 1$  cut along a radius from 0 to 1 on a surface  $S$  with function  $(1/k)p(R^k)$ ; hence  $\{f(z^k)\}^{1/k}$  (which maps  $|z| < 1$  on  $S$  covered  $k$ -times) has for function

$$k \cdot (1/k)p(R^k) = p(R^k).$$

Finally

$$\int_0^{R_1} p_k(R) d(\pi R^2) = \int_0^{R_1} p(R^k) d(\pi R^2) = \int_0^{R_1^k} p(R) d(\pi R^{2/k}) \leq p \pi R^2,$$

by the mean  $p$ -valency of  $f$  and Lemma 1. This proves the first half of the theorem.

As for the second half, let  $p_1(R)$ ,  $p_k(R)$  correspond respectively to  $f_1 = \{f_k(z^{1/k})\}^k$ ,  $f_k(z)$ . Then

$$p_1(R) = p_k(R^{1/k}).$$

An argument similar to that given above to prove the negative part of Lemma 1 now shows that there exist mean  $p$ -valent functions  $f_k$  and arbitrarily large  $R_1$  for which

<sup>(7)</sup> Since  $x^s$  increases practically from 0 to 1 in an arbitrarily small neighborhood of  $x=1$  when  $s$  is large.

$$\int_0^{R_1} p_1(R) d(\pi R^2) = \int_0^{R_1} p_k(R^{1/k}) d(\pi R^2) = \int_0^{R_1^{1/k}} p_k(R) d(\pi R^{2k}) > p \pi R_1^2,$$

if  $k > 1$ . If, however,  $p_k(R) \leq p$ , then  $p_1(R) \leq p$ , and in this case (in particular if  $f_k$  is  $p$ -valent)  $f_1$  is mean  $p$ -valent.

5. After these preliminaries we now study the rate of growth of mean  $p$ -valent functions. The method depends on the distortion theory of Ahlfors<sup>(8)</sup>, a theory which has already been applied by Cartwright<sup>(9)</sup> to obtain an upper bound of  $M(r, f)$  (the maximum modulus of  $f(z)$  on  $|z| = r$ ) for  $p$ -valent functions. By  $K(\alpha, \beta, \dots)$  we denote a positive number depending on the parameters shown explicitly. If it is clear on what parameters  $K$  depends, as often happens, we simply write  $K$ .  $K$ 's will not necessarily be the same in different contexts.

It is convenient to suppose first that  $f(z)$  is regular for  $|z| \leq 1$ . We write  $w_0 = f(0)$ . Let  $C(R)$  be the circumference  $|w| = R$  in the  $w$ -plane, and let  $E(R) = W \times C(R)$ , the set of points common to  $W$  and  $C(R)$  (so that  $mE(R) = 2\pi R p(R)$ ). Two points of  $E(R)$  are considered distinct if they correspond to distinct sheets of  $W$ , even though they have the same projection on the complex  $w$ -plane.  $E(R)$  consists of a finite set of arcs  $\{I_\nu(R)\}$ <sup>(10)</sup>, ( $\nu = 1, 2, \dots, N$ ), where  $N$  depends on  $R$  (and  $f$ ). For fixed  $R_1$  let  $r_\nu(R_1)$  be the value of  $r$  for which  $B(r)$  (the transform of  $|z| = r$  by  $f$ ) just touches  $I_\nu(R_1)$  (for the first time). If  $|w_0| < R < R_1$ , at least one arc of  $E(R)$  separates  $I_\nu(R_1)$  from  $w_0$ ; if more than one, let  $I_\nu(R)$  be the first which is met in describing a continuous curve lying in  $W$  and connecting  $w_0$  with a point of  $I_\nu(R_1)$ . Let  $mI_\nu(R) = \Theta_\nu(R)$ .

**THEOREM 3.** Suppose that  $0 < r < 1$ , and that  $R_1 > M(r_0, f)$ . Then

$$(5.1) \quad 2\pi \int_{M(r_0, f)}^{R_1} \frac{dR}{\Theta_\nu(R)} \leq \log \frac{1}{(1 - r_\nu(R_1))^2} + K(r_0), \quad (\nu = 1, 2, \dots, N(R_1)).$$

Take  $R_1 = M(r, f)$ , and let  $I_{\nu(r)}$  be any one of the intervals  $\{I_\nu(R_1)\}$  which is touched by  $B(r)$  (there is at least one). Then, if  $r > r_0$ , we have by Theorem 2 (with  $R_1 = M(r, f)$ ,  $\nu = \nu(r)$ )

$$(5.2) \quad 2\pi \int_{M(r_0, f)}^{M(r, f)} \frac{dR}{\Theta_{\nu(r)}(R)} \leq \log \frac{1}{(1 - r)^2} + K(r_0).$$

This formula has been proved in effect by Cartwright [3]. I omit the proof of the more general formula (5.1) since no essentially new ideas are involved.

## 6. Let

<sup>(8)</sup> Ahlfors [1].

<sup>(9)</sup> Cartwright [3].

<sup>(10)</sup>  $C(R)$  may not cut  $B$  (the transform of  $|z| = 1$  and the boundary of  $W$ ), in which case each interval of  $E(R)$  is the whole of  $C(R)$ , and the number of intervals is the number of sheets cut by  $C(R)$  (zero for large  $R$ ).



$$f_k(z) = a_1 z + a_{k+1} z^{k+1} + \dots$$

We deduce the following theorem from Theorem 3:

**THEOREM 4.** *If  $f_k(z)$  is mean  $p$ -valent and  $M(r, f_k)$  is the maximum modulus of  $f_k$  on the circle  $|z| = r$ , then*

$$(6.1) \quad M(r, f_k) \leq K(p, k) \mu_{[p/k]} (1-r)^{-2p/k},$$

where

$$\mu_{[p/k]} = \max (|a_1|, \dots, |a_{[p/k]}|).$$

Theorem 4 was stated without proof in  $V_1$ . It is known for  $p$ -valent functions, the case  $k=1$  having been proved by Cartwright (loc. cit.); and the general case is an easy deduction from the case  $k=1$  when  $f_k$  is  $p$ -valent<sup>(11)</sup>. By combining Theorem 4 above with Theorem 3 of  $V_1$  we obtain the following theorem (also stated without proof in  $V_1$ ):

**THEOREM 5.** *If  $f_k(z)$  is mean  $p$ -valent, then*

$$|a_n| \leq K(p, k) \mu_{[p/k]} n^{2p/k-1}$$

provided  $p > \frac{1}{2}k$ .

Theorem 5 for  $p$ -valent functions was proved in  $V_1$ <sup>(12)</sup>. The restriction that  $p > \frac{1}{2}k$  is necessary. In fact, if  $n \geq 1$ , take

$$a_{nk+1} = \begin{cases} \frac{1}{v(\lambda_r)^{1/2}} z^{nk+1}, & \text{if } |(nk+1) - \lambda_r| \leq k/2 \text{ and } \lambda_r \geq nk+1, \\ 0, & \text{otherwise,} \end{cases}$$

where  $(\lambda_r)$  is a rapidly increasing sequence, and take  $a_1 = 1$ . We suppose that the  $\lambda_r$  satisfy the inequality

$$\sum_{r=1}^{\infty} \frac{1}{v(\lambda_r)^{1/2}} \leq 1 - (\pi/6)^{1/2}.$$

Then  $f_k(z)$  is zero only at the origin, and each point of the circle  $|w| < (\pi/6)^{1/2}$  is covered by  $W_k$  (the transform of  $|z| < 1$  by  $f_k$ ) once and only once. Since the area of  $W_k$  is less than or equal to  $\pi \sum_{r=1}^{\infty} 1/v^2 = \pi^2/6$ , we see that  $W(R) \leq \pi^2 R^2$  for  $R > 0$ , so that  $f_k$  is mean  $\pi$ -valent. On the other hand, given any function  $\psi(n)$  tending steadily to 0 as  $n \rightarrow \infty$ , we can choose the  $\lambda_r$  such that  $|a_n| > \psi(n)/n^{1/2}$  for an infinity of  $n$ . This *Gegenbeispiel* in modified form was suggested to me by Professor J. E. Littlewood.

<sup>(11)</sup> For then the function  $f_1(z) = \{f_k(z^{1/k})\}^k$  is  $p$ -valent. This line of argument is not possible here (see Theorem 2).

<sup>(12)</sup> The theorem for  $p$ -valent functions was known subject to certain restrictions on  $f_k$ ; in  $V_1$  these restrictions were removed.

7. We now prove Theorem 4. We define  $\Theta_r(\rho, R)$ , the function of Theorem 3, in terms of  $f_k(\rho, z)$ , where  $\rho < 1$ . Then we define

$$\Theta_r(R) = \lim_{\rho \rightarrow 1} \Theta_r(\rho, R) \leq \lim_{\rho \rightarrow 1} 2\pi R p(\rho, R) = 2\pi R p(R),$$

since  $f_k$  is mean  $p$ -valent.  $\Theta_r(R)$  is thus an integrable function of  $R$  (over any finite interval). Now let  $W_k$  be the transform of  $|z| < 1$  by  $f_k$ . Since rotation of  $W_k$  about the origin through an angle  $2\pi/k$  transforms  $W_k$  into itself, we see that if  $T_0$  is any "tube" of  $W_k$  extending to  $\infty$ , there are  $(k-1)$  other tubes  $T_\nu$ , ( $\nu = 1, 2, \dots, (k-1)$ ), each identical to  $T_0$ . If, therefore,  $\Theta_r(R)$  is the width of  $T$ , measured on  $C(R)$ , we have

$$\sum_{r=0}^k \Theta_r(R) = k\Theta_0(R) \leq mE(R) = 2\pi p(R),$$

and so

$$\Theta_0(R) \leq \frac{2\pi}{k} R p(R).$$

Hence

$$(7.1) \quad 2\pi \int_{M(r_0)}^{M(r)} \frac{dR}{\Theta_{r(r)}(R)} \geq k \int_{M(r_0)}^{M(r)} \frac{1}{p(R)} \frac{dR}{R} = k \int_{\log M(r_0)}^{\log M(r)} \frac{dR}{p(e^R)} \\ \geq k \frac{(\log M(r) - \log M(r_0))^2}{\int_{\log M(r_0)}^{\log M(r)} p(e^R) dR},$$

since (writing  $\psi(R) = 1/p(e^R)$ )

$$(b-a)^2 = \left( \int_a^b \psi^{1/2} \psi^{-1/2} dR \right)^2 \leq \int_a^b \psi dR \int_a^b \psi^{-1} dR$$

by Schwarz's inequality. But

$$(7.2) \quad \int_{R_1}^{R_2} p(e^R) dR = \int_{R_1}^{R_2} p(R) \frac{dR}{R} \\ = \left[ \frac{1}{R} \int_{R_1}^R p(R) dR \right]_{R_1}^{R_2} + \int_{R_1}^{R_2} \left( \int_0^R p(R) dR \right) \frac{1}{R^2} dR \\ \leq p + p(R_2 - R_1)$$

by the hypothesis of mean valency  $p$  and Lemma 1 (with  $s_1=2$ ,  $s_2=1$ ). Substituting from (7.2) (with  $R_1 = \log M(r_0)$ ,  $R_2 = \log M(r)$ ) in (7.1) and using (5.2), we have

$$(7.3) \quad \log M(r) \leq \frac{p}{k} \log \frac{1}{(1-r)^2} + K(p, k, r_0) + \log M(r_0).$$

Theorem 4 will follow at once from (7.3) (with  $r_0 = \frac{1}{2}$ , say) if

$$(7.4) \quad M(r_0, f) < K(p, k, r_0) \mu_{[p/k]}.$$

To prove (7.4) it is sufficient to show that the family of mean  $p$ -valent functions  $f_k$  is quasi-normal of order  $[p/k]$  at most, and this follows from the definition of mean valency  $p$  and the form of  $f_k$  (13). This completes the proof of Theorem 4.

8. The full strength of the hypothesis of mean valency  $p$  is not used in Theorem 4; all that is used is (7.2), and this in the form

$$(1) \quad \int_{R_1}^{R_2} p(e^R) dR \leq p/s + p(R_2 - R_1),$$

where  $s > 0$ , is implied by (8). The hypothesis (1) is, in fact, sufficient for the truth of all theorems proved in this paper. Furthermore, only the properties of  $W$  in the neighborhood of  $\infty$  are relevant. For example, if  $W(R) \leq p\pi R^2$  only for  $R > R_0 > 0$ , then

$$M(r, f) = O(1 - r)^{-2p},$$

where the constant implied in the  $O$  depends on  $R_0, f$ , and  $p$ . More generally if

$$(8.1) \quad \limsup_{R \rightarrow \infty} \frac{W(R)}{\pi R^2} \leq p,$$

then, for every  $\epsilon > 0$ ,

$$M(r, f) = O(1 - r)^{-2p-\epsilon}.$$

Moreover, if (8.1) is satisfied with  $p=0$ , then

$$M(r, f) = O(1 - r)^{-\epsilon}.$$

We thus obtain, in particular, the striking result that a schlicht function which fills only an infinitesimal part of the  $w$ -plane is of infinitesimal order.

9. We shall say that a set of points in a domain  $D$  is a path  $P$  if it is a Jordan curve. If the equation of  $P$  is

$$P(t) = x(t) + iy(t),$$

where  $t$  varies from 0 to 1, and if, given  $\epsilon$ ,

$$|P(t) - a| < \epsilon$$

for  $t_0(\epsilon) < t < 1$ , then we say  $a$  is an end of  $P$ , or that  $P$  converges to the point  $a$ . A path in  $|z| < 1$  with end  $e^{i\theta}$  will be denoted by  $P(\theta)$ .

(13) See Montel [5, p. 73]. The test given there for quasi-normality  $[p/k]$  is satisfied if applied to the functions  $f_k(z^{1/k})$ , which are regular in the unit circle slit along a radius, and this implies that the family  $f_k(z)$  is quasi-normal of order  $[p/k]$ .

THEOREM 6. Suppose that  $f(z)$  is *p.m.v.* and that  $E_\theta$  is a set of distinct points. If to each point  $\theta$  of  $E_\theta$  there corresponds at least one path  $P(\theta)$  for which

$$(9.1) \quad \liminf_{P_\theta} (1-r)^{\alpha(\theta)} |f(z)| > 0,$$

then

$$(9.2) \quad \sum_{E_\theta} \alpha(\theta) \leq 2p.$$

It is sufficient to prove the theorem for an enumerable set  $(\theta_r)$ <sup>(14)</sup>. It is then enough to show that

$$(9.3) \quad \sum_{r=1}^n \alpha_r \leq 2p,$$

where  $\alpha_r = \alpha(\theta_r) > 0$ . Under these circumstances there correspond to  $R > R_0(n, f)$ ,  $n$  arcs  $I_\nu(R)$ ,  $(1 \leq \nu \leq n)$ , such that the transform of  $I_\nu(R)$  by  $z = f^{-1}(w)$  is a cross section<sup>(15)</sup>  $\gamma_\nu(R)$  of the unit circle separating the point  $e^{i\theta_r}$  from the origin and converging to  $e^{i\theta_r}$  as  $R \rightarrow \infty$ <sup>(16)</sup>. Let  $R_\nu(r)$  be the largest  $R$  for which  $\gamma_\nu(R)$  has points in common with the circle  $|z| = r$ , and write

$$mI_\nu(R) = \Theta_\nu(R) = 2\pi R \Xi_\nu(R).$$

Then

$$2\pi \int_K^{R_\nu(r)} \frac{dR}{\Theta_\nu(R)} = \int_{K_1}^{\log R_\nu(r)} \frac{dR}{\Xi_\nu(e^R)} \geq \frac{(\log R_\nu(r) - R_1)^2}{\int_{R_1}^{\log R_\nu(r)} \Xi_\nu(e^R) dR},$$

as in the proof of Theorem 4. That is to say,

$$\int_{K_1}^{\log R_\nu(r)} \Xi_\nu(e^R) dR \geq \frac{(\log R_\nu(r) - R_1)^2}{2\pi \int_K^{R_\nu(r)} dR / \Theta_\nu(R)} \geq \frac{(\log R_\nu(r) - K_1)^2}{\log 1/(1-r)^2 + K}$$

by Theorem 3,

$$\geq \frac{1}{2} \alpha_r \log R_\nu(r) + o(\log R_\nu(r)),$$

by the hypothesis (9.1). This inequality may be written in the form

$$\frac{1}{2} R \alpha_r \leq \int_K^R \Xi_\nu(e^R) dR + o(R).$$

Summing over  $\nu$  from 1 to  $n$ , we obtain

<sup>(14)</sup> But even a schlicht function may tend to  $\infty$  at a non-enumerable set of discrete points  $e^{i\theta}$ .

<sup>(15)</sup> By a cross section of a domain  $D$  we mean a path lying in  $D$  (except for its end-points) and connecting two distinct boundary points of  $D$ .

<sup>(16)</sup> That is, given  $\epsilon$ ,  $\gamma_\nu(R)$  lies in a circle of radius  $\epsilon$  and center  $e^{i\theta_r}$  if  $R > R_0(\epsilon)$ . The statement is intuitive, and in any case is covered by familiar arguments.

$$\frac{1}{2}R \sum_{r=1}^n \alpha_r \leq \int_K^R \sum_{r=1}^n \Xi_r(e^R) dR + o(R) \leq \int_K^R p(e^R) dR + o(R),$$

since  $\sum_{r=1}^n \Xi_r(R) \leq p(R)$ ,

$$\leq pR + o(R),$$

by the hypothesis of mean valency  $p$  (see (7.2)). Dividing by  $R$  and letting  $R \rightarrow \infty$ , we obtain (9.3), and (since  $n$  is arbitrary) this proves the theorem.

10. THEOREM 7. Suppose  $f(z)$  is *p.m.v.* and that

$$(10.1) \quad f(z) = O(1)$$

on some path  $P_1(\theta_0)$ . Then on any path  $P(\theta_0)$

$$(10.2) \quad \limsup (1-r)^{2p} |f(z)| = 0.$$

We suppose there is an infinite sequence of points,  $(z_n)$  say, tending to  $e^{i\theta_0}$  and a number  $K > 0$  such that  $|f(z_n)| > |f(z_{n-1})|$ , and

$$(10.3) \quad |f(z_n)| > K(1-r_n)^{-2p}, \quad |z_n| = r_n;$$

we argue by *reductio ad absurdum*.

Suppose first that there exists an arbitrarily large  $R$  such that the transform of  $E(R)$  by  $z = f^{-1}(w)$  contains an infinity of nonoverlapping cross sections  $\gamma_r(R)$  of  $|z| < 1$  converging to  $e^{i\theta_0}$  as  $r \rightarrow \infty$  <sup>(17)</sup>, and that each  $\gamma_r$  separates at least one point  $z_n$  from the origin. Changing the numeration (if necessary) we may suppose that  $\gamma_r(R)$  separates  $z_r$  from  $z = 0$ . Let  $I_r(R)$  be the transform by  $f(z)$  of  $\gamma_r(R)$ , and write  $mI_r(R) = \Theta_r(R) = 2\pi R \Xi_r(R)$ ,  $R_r = |f(z_r)|$ . Then

$$(10.4) \quad p \log R_r \leq \frac{(\log R_r - K)^2}{2\pi \int_K^{R_r} dR / \Theta_r(R)} + O(1).$$

Otherwise

$$\log R_r < 2\pi p \int_K^{R_r} \frac{dR}{\Theta_r(R)} + O(1) < \log \frac{1}{(1-r_r)^2} + O(1)$$

by Theorem 3, and this contradicts the hypothesis (10.3). But (as in the proof of Theorem 6)

$$\frac{(\log R_r - K)^2}{2\pi \int_K^{R_r} dR / \Theta_r(R)} \leq \int_K^{\log R_r} \Xi_r(e^R) dR,$$

and so, substituting in (10.4),

<sup>(17)</sup> But no  $\gamma_r$  separates  $e^{i\theta_0}$  from the origin.



$$(10.5) \quad p \log R, \leq \int_K^{\log R} \Xi_r(e^R) dR + O(1).$$

Finally, since  $\gamma_{r-1}$  and  $\gamma_r$  are nonoverlapping, we see that  $\Xi_{r-1}(R)$  and  $\Xi_r(R)$  are distinct for all (sufficiently) large  $R$ , and so

$$(10.6) \quad \int_K^{\log R} \Xi_r(e^R) dR \leq \int_K^{\log R} p(e^R) dR - \int_K^{\log R} \Xi_{r-1}(e^R) dR \\ \leq p \log R - p \log R_{r-1} + O(1)$$

by the hypothesis of mean valency  $p$  and (10.5) for  $\nu-1$ . (10.5) and (10.6) give a contradiction if  $\nu > \nu_0$ , and so the infinity of nonoverlapping cross sections with the properties stated cannot exist. The alternative is that for  $R > R_0$  one cross section,  $\gamma(R)$  say, separates all but a finite number of  $(z_r)$  from  $z=0$ .

Now we can find a number  $R_1$  such that, for  $R > R_1$ ,  $\gamma(R)$  does not separate  $e^{i\theta_0}$  from  $z=0$ . Otherwise there would exist no path  $P_1(\theta_0)$  on which  $f=O(1)$ , contrary to the hypothesis of the theorem. Since, on the other hand,  $\gamma(R)$  separates all but a finite number of the  $(z_r)$  from  $z=0$ , we see that, for  $R > R_1$ ,  $\gamma(R)$  has  $e^{i\theta_0}$  as one end-point. This, I say, is impossible<sup>(18)</sup>. In fact, suppose  $R_1 < R_2 < R_3$ , and connect  $\gamma(R_2)$  with  $\gamma(R_3)$  by a simple analytic curve lying in  $|z| < 1$ . Let  $q_1$  be the last intersection of this curve with  $\gamma(R_2)$ ,  $q_2$  the first intersection with  $\gamma(R_3)$ . Then the portion  $P$  of the curve connecting  $q_1$  with  $q_2$  lies in a sub-domain  $D$  of  $|z| < 1$  (bounded by  $\gamma(R_2)$ ,  $\gamma(R_3)$ , and points of  $|z|=1$ ), and divides  $D$  into two domains. Let  $D_1$  be the domain bounded by  $\gamma(R_2)$ ,  $\gamma(R_3)$ ,  $P$ , and  $e^{i\theta_0}$ ; and let  $W_1$  be the transform of  $D_1$ ,  $\Pi$  the transform of  $P$ , by  $f$ . Suppose  $R_2 < R < R_3$ , and let  $I(R)$  be the first cross section of  $W_1$  on  $C(R)$  which is met in describing a continuous curve from  $E(R_2)$  to  $E(R_3)$  in  $W_1$ . We write  $\Theta(R) = mI(R)$ . Then, by the hypothesis of mean valency  $p$ ,

$$\int_{R_1}^{R_3} mI(R) dR \leq p\pi R_2^2.$$

Hence, if  $K = 2p\pi R_2^2/(R_3 - R_1)$ , and  $E$  is the set of values of  $R$  in the interval  $R_1 < R < R_3$  for which

$$mI(R) > K,$$

then

$$(10.7) \quad mE < \frac{1}{2}(R_3 - R_1).$$

Next, we define

$$J(R) = \begin{cases} I(R), & \text{if } mI(R) \leq K, \\ \text{a portion of } I(R) \text{ of length } K \text{ measured from } \Pi & \text{if } mI(R) > K. \end{cases}$$

<sup>(18)</sup> For finitely mean valent functions, but not for infinitely mean valent functions.

Let  $W_2$  be one of the sub-domains of  $W_1$  swept out by  $J(R)$  as  $R$  varies from  $R_2$  to  $R_3$ , which contains, as part of its boundary, a set  $A$  of boundary points of  $W$  of positive measure. Such a sub-domain exists by (10.7). Further,  $W_2$  is plainly a finitely valent domain; and every point of its boundary is accessible (by the definition of accessibility). We map  $W_2$  on a sub-domain  $D_2$  of  $|z| < 1$  by  $f^{-1}$ , the set  $A$  corresponding to the boundary point  $e^{i\theta_0}$ . This contradicts well known theorems on the correspondence of boundaries<sup>(19)</sup> and proves our statement.

11. The conclusion (9.2) of Theorem 6 is a best possible one when  $p$  is integral, as shown by the  $p$ -valent function

$$f_n(z) = \frac{z^p}{(1 + z^n)^{2p/n}}.$$

On the other hand, the hypothesis (9.1) cannot be relaxed to the extent of replacing  $\liminf$  by  $\limsup$ . We have in fact

THEOREM 8. If  $\psi(r)$  is any real function of  $r$  satisfying

$$(11.1) \quad (1 - r)^2 = o(\psi(r)),$$

then there is a function  $f(z)$  regular and schlicht in  $|z| < 1$  such that, for at least one path  $P(\theta_r)$ ,

$$(11.2) \quad \limsup_{P(\theta_r)} \psi(r) |f(z)| > 0$$

at an enumerable infinity of discrete points  $(\theta_r)$ .

The following theorem shows that Theorem 7 is best possible.

THEOREM 9. Suppose  $\psi(r)$  satisfies (11.1). Then there is a schlicht function  $f(z)$  such that the radial limit,  $\lim_{r \rightarrow 1} f(re^{i\theta})$ , exists everywhere and is finite, but

$$(11.3) \quad \limsup \psi(r) |f(z)| > 0$$

on at least one path  $P(\theta_0)$ .

The function whose existence is asserted in Theorem 9 is simpler and we discuss it first. We take  $f(z)$  to be the function which maps  $|z| < 1$  on the simply-connected domain  $W$  shown in Fig. 2, with  $f(0) = 0$  and  $f'(0)$  real and positive (so that  $f$  is uniquely defined by  $W$ ).  $W$  consists of the whole  $w$ -plane slit along an infinity of concentric circles of radii  $R_\nu$ , ( $\nu = 1, 2, \dots$ ), each annular region  $(R_\nu, R_{\nu+1})$  being connected by a "thin tube" to the interior of the circle of radius  $R_1$ . Every point of the boundary  $B$  of  $W$  is accessible except points on the line extending from  $\omega$  to  $\infty$ . The line from  $\omega$  to  $\infty$  is an infinite prime-end with the single accessible nuclear point (Hauptpunkt)

<sup>(19)</sup> See, for example, Carathéodory [2].



$\gamma_\nu(R)$  converges to  $e^{i\theta_0}$  as  $\nu \rightarrow \infty$ ,  $z_\nu$  tends to the same point (as  $\nu \rightarrow \infty$ ) and

$$\psi(r_\nu) |f(z_\nu)| = \psi(r_\nu) \frac{R_\nu + R_{\nu+1}}{2} > K$$

by (11.4).

If, having chosen  $R_\nu$ , we can choose  $R_{\nu+1}$  such that (11.4) is satisfied by the function  $f_\nu(z)$  which maps (with  $f_\nu(0)=0$ ,  $f'_\nu(0)>0$ ) the circle  $|z|<1$  on the sub-domain  $W_\nu$  of  $W$  shown in Fig. 3, it will follow by the subordination principle<sup>(21)</sup> that (11.4) is *a fortiori* satisfied by  $f$  uniformly in  $\nu$ .

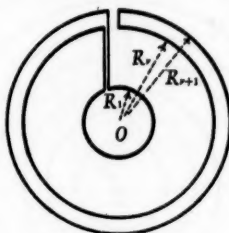


FIG. 3

To show that, for suitable choice of  $R_{\nu+1}$ ,  $f_\nu$  satisfies (11.4), we cut  $W_\nu$  along a radius from 0 to a point  $w$  of its boundary on  $|w|=R_\nu$ , and map the resulting domain by means of  $s_\nu(z)=\sigma+i\tau=\log w$  on a strip  $S_\nu$ . Now for a parallel strip  $U$  defined by

$$\zeta(z) = \xi + i\eta, \quad \xi_0 < \xi < \xi_1, \quad |\eta| < a\pi,$$

we have

$$(11.5) \quad M(r, \zeta) \geq \log \frac{1}{(1-r)^{2a}} + K(\xi_0)$$

if  $\xi_0 + A < M(r, \zeta) < \xi_1 - A$ . But for suitable  $z_0$  (depending on  $R$ ), the function  $s_\nu(h(z))$ , where

$$h(z) = \frac{z - z_0}{z\bar{z}_0 - 1},$$

is superordinate to a  $U$  with  $\xi_0 = \log R_\nu$ ,  $\xi_1 = \log R_{\nu+1}$  and  $a = 1 - 1/\xi_1$  (since we may make the angular spread of the annular region of  $W_\nu$  as near to  $2\pi$  as we please). Hence

$$M(r, s_\nu) \geq M(r, s_\nu(h)) - K(R_\nu) \geq M(r, \zeta) - K(R_\nu),$$

by subordination. If  $K_1(R_\nu) < M(r, \zeta) < \log R_{\nu+1} - K_1$ , this is not less than

<sup>(21)</sup> See Littlewood [6].

$$\log \frac{1}{(1-r)^2} - \frac{1}{\xi_1} \log \frac{1}{1-r} - K(R_r),$$

by (11.5); and is greater than or equal to

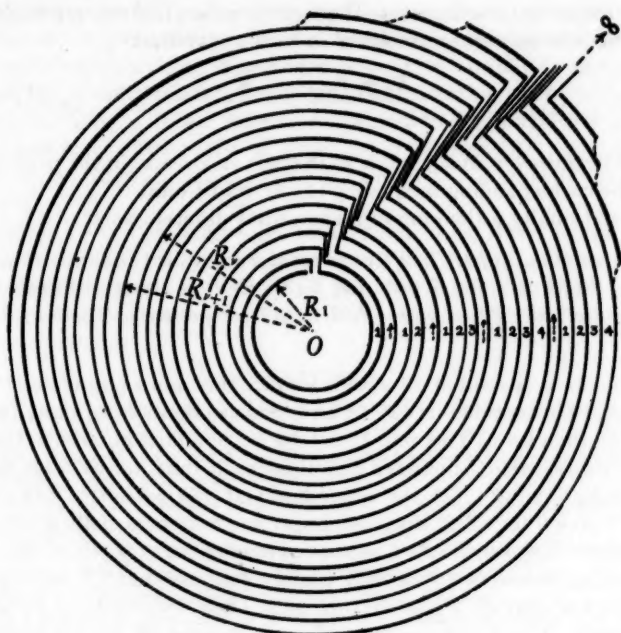


FIG. 4

$$\log \frac{1}{(1-r)^2} - K(R_r) \geq \log \frac{K}{\psi(r)}$$

if  $r$  (and so  $R_{r+1}$ ) is large enough. In particular,

$$\log \frac{R_r + R_{r+1}}{2} = M(r_r, s_r) \geq \log \frac{K}{\psi(r_r)}$$

if  $R_{r+1} > R_0(R_r)$ . We can thus choose  $R_{r+1}$  so that  $f_r$  satisfies (11.4), and this completes the proof of Theorem 7.

12. In Theorem 8 let  $f(z)$  be the function which maps  $|z| < 1$  on the  $w$ -plane slit as shown in Fig. 4. The domain consists of an infinity of "tubes" (numbered as shown) connecting the circle  $|w| \leq R_1$  with  $\infty$ . If we write

$$N(n) = \sum_{\mu=1}^n \mu,$$



then the  $\nu$ th tube has an angular spread of nearly  $2\pi$  over the "long" intervals (of  $R$ ):

$$(12.1) \quad R_{N(n)+\nu-1} < R < R_{N(n)+\nu} \quad (n \geq \nu).$$

Using the same notation as in Theorem 9, we see it is enough to show that the radii may be chosen successively in such a way that

$$\frac{R_\nu + R_{\nu+1}}{2} > \frac{K}{\psi(r_\nu)} \quad (\nu = 1, 2, \dots).$$

The proof is, however, now similar to that given already in the preceding section for the corresponding inequality (11.4), and I omit it.

13. I add finally a theorem of a somewhat different sort:

**THEOREM 10.** Suppose that  $f(z)$  is regular in  $|z| < 1$ , and satisfies the condition that  $W(R) < \infty$ ,  $0 \leq R < \infty$ . Let  $E_1(\theta)$ ,  $E_2(\theta)$  be the sets of limit points as  $f(z)$  tends to  $e^{i\theta}$  along two paths  $P_1(\theta)$ ,  $P_2(\theta)$  respectively. Then  $E_1(\theta) \times E_2(\theta) \neq \emptyset$  <sup>(22)</sup>.

This theorem is related to a well known theorem of Lindelöf <sup>(23)</sup> which states that, if  $f$  is bounded in  $|z| < 1$  and tends to limits  $l_1$ ,  $l_2$ , along two paths  $P_1(\theta)$ ,  $P_2(\theta)$ , then  $l_1 = l_2$ . Theorem 10 is false for bounded functions; there exist (infinitely mean valent) bounded functions such that, for at least one point  $e^{i\theta}$ ,  $E_1(\theta) \times E_2(\theta) = \emptyset$  <sup>(24)</sup>. On the other hand, if  $F(\theta) = E_1(\theta) \times E_2(\theta)$ , the hypothesis " $F(\theta) \neq \emptyset$  for all  $\theta$ " does not imply the finiteness of  $W(R)$  <sup>(25)</sup>, so that the conditions  $F(\theta) \neq \emptyset$ ,  $W(R) < \infty$  are not equivalent.

In proving Theorem 10 we may plainly suppose that  $|f|$  is bounded on  $P_1(\theta)$ ,  $P_2(\theta)$  and that  $P_1$ ,  $P_2$  do not intersect. Then, joining  $P_1$  to  $P_2$  by a path  $Q$  lying inside  $|z| < 1$ , we can map the sub-domain of  $|z| < 1$  bounded by  $P_1$ ,  $P_2$ , and  $Q$ , onto the unit circle, the paths  $P_1$ ,  $P_2$  being transformed into two arcs,  $\Pi_1$ ,  $\Pi_2$ , abutting at a point  $e^{i\theta}$ . Let  $L_1$ ,  $L_2$  be the transforms of  $\Pi_1$ ,  $\Pi_2$  by  $f$ , and let  $\Lambda_1$ ,  $\Lambda_2$  be the projections of  $L_1$ ,  $L_2$  on the  $w$ -plane. We suppose  $E_1 \times E_2 = \emptyset$ , and argue by *reductio ad absurdum*.

If  $E_1 \times E_2 = \emptyset$ , there exist two positive numbers  $\delta$  and  $r_1$  such that the portions of  $\Lambda_1$  and  $\Lambda_2$  corresponding to the arc of  $|z| = 1$  which lies inside a circle of radius  $r_1$  and center  $e^{i\theta}$  are separated by a distance  $\delta$ . Let  $c(r)$  be that arc of the circle of radius  $r$  and center  $e^{i\theta}$  which lies in  $|z| < 1$ , and let  $\Gamma(r)$  be

<sup>(22)</sup> That is,  $E_1$  and  $E_2$  contain a common point (which may be  $\infty$ ).

<sup>(23)</sup> Lindelöf [4].

<sup>(24)</sup> An example is the function  $f$  which maps the unit circle on the circle  $|w| \leq R$  covered infinitely many times, with winding point at  $w=0$ . There is then one point  $e^{i\theta}$ , and two paths  $P_1$ ,  $P_2$  converging to it, such that the transforms of  $P_1$  and  $P_2$  by  $f$  are concentric circles.

<sup>(25)</sup> In fact, if  $f$  maps the unit circle on a Riemann domain bounded by a "spiral" with asymptotic point  $w=0$ , then  $f$  tends to a limit on every path  $P(\theta)$ . By coiling the spiral sufficiently loosely, the sum of the areas bounded by successive loops can be made infinite.

that portion of the transform of  $c(r)$  which connects  $\Lambda_1$  to  $\Lambda_2$ . Let  $W_{r_1}$  be the simply connected domain bounded by  $\Gamma(r_1)$ ,  $\dots$ ,  $\Gamma(r_1)$ , and subsets  $B_1(r_1)$ ,  $B_2(r_1)$  of the boundary continua  $\Lambda_1$  and  $\Lambda_2$ . Now I say no point  $w$  is covered by  $W_{r_1}$  more than a finite number of times. For, by the construction of  $W_{r_1}$ , the boundary curves  $B_1(r_1)$ ,  $B_2(r_1)$  are both simple, and  $\Gamma(r_1)$  (the transform of a portion of  $c(r_1)$ ) is analytic. Hence, if a point  $w$  were covered an infinity of times, some neighborhood of  $w$  would be covered an infinity of times, and the area of  $W_{r_1}$  would therefore be infinite, contradicting the hypothesis that  $W(R) < \infty$ , for finite  $R$  (since  $W_{r_1}$  is a finite domain). Similarly, if  $W_{r_1}'$  is an interior domain, the boundary of which is at distance  $\delta/4$ , say, from the boundary of  $W_{r_1}$ , then the valency of points of  $W_{r_1}'$  is uniformly bounded by a number  $K$ .

Finally, as  $r \rightarrow 0$ ,  $\Gamma(r)$  converges to an "end"  $\xi$  of  $W(r_1)$  in the sense of Carathéodory<sup>(24)</sup>. Since  $W(r_1)$  is bounded and  $\Lambda(r_1)$ ,  $\Lambda(r_2)$  are separated by a distance  $\delta/2$ ,  $\Gamma(r)$  cannot converge to a point or to  $\infty$ . Therefore,  $\xi$  is not a prime-end, and so cannot correspond to a *single* point. This is a contradiction and proves the theorem.

#### REFERENCES

1. L. Ahlfors, *Untersuchungen zur Theorie der konformen Abbildung und der ganzen Funktionen*, Acta Academiae Scientiarum Fennicae, vol. 1 (1930), no. 7.
2. C. Carathéodory, *Über die Begrenzung einfach zusammenhängender Gebiete*, Mathematische Annalen, vol. 73 (1912), pp. 323-370.
3. M. L. Cartwright, *Some inequalities in the theory of functions*, Mathematische Annalen, vol. 111 (1935), pp. 98-118.
4. E. Lindelöf, *Sur un principe général de l'analyse et ses applications à la théorie de la représentation conforme*, Acta Academiae Scientiarum Fennicae, vol. 46 (1915).
5. P. Montel, *Leçons sur les Familles Normales*, Paris, 1927.
6. J. E. Littlewood, *On inequalities in the theory of functions*, Proceedings of the London Mathematical Society, (2), vol. 23 (1925), pp. 481-519.

<sup>(24)</sup> See Carathéodory [2]. Carathéodory develops his theory only for schlicht functions; here we require its extension to finitely valent functions. The extension is, however, trivial.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,  
CAMBRIDGE, MASS.

# INTEGRAL SETS OF QUATERNION ALGEBRAS OVER A FUNCTION FIELD

BY

LEONARD TORNHEIM

1. **Introduction.** The theory of rational quaternion algebras suggests a corresponding theory for quaternion algebras over a rational function field  $F(z)$ . We can anticipate points of close analogy because of many similarities between the set of all rational integers and the set of polynomials  $F[z]$ . We may also expect results peculiar to each of these theories traceable to certain fundamental differences between these two integral domains.

We find a basis for every integral set  $S$  of  $Q$  after suitably normalizing a basis of  $Q$ . When  $F$  is the field of all real numbers, canonical bases for both  $Q$  and  $S$  are obtained. We discuss properties of  $Q$  which make  $S$  be a principal ideal ring or not. Conditions are provided for a quantity in  $F[z]$  to generate a prime ideal in  $S$ . Throughout applications are made in the cases for which  $F$  is either a real number field or a finite field.

2. **Integral sets of  $Q$  with characteristic not two.** When  $F$  has characteristic not two, a quaternion algebra<sup>(1)</sup>  $Q$  has a basis  $1, i, j, ij$  with  $i^2 = \tau, j^2 = \sigma, ij = -ji$ . The basis can be chosen as *normalized*<sup>(2)</sup>; that is,  $\sigma, \tau$  lie in  $F[z]$ , are relatively prime, and contain no square factors. If  $F$  is a Hilbert irreducibility field<sup>(3)</sup>, it is possible in addition to take  $\tau$  a prime (i.e., irreducible) in  $F[z]$ .

Integral sets<sup>(4)</sup>  $S$  of  $Q$  are defined by the usual four properties R, C, U, and M. We obtain a basis for an integral set  $S$  of  $Q$  in

**THEOREM<sup>(5)</sup> 1.** *Let  $Q$  have a normalized basis. Let  $\tau'$  be the product of all prime factors of  $\tau$  for which  $\sigma$  is a quadratic residue, and  $\tau'' = \tau/\tau'$  be monic<sup>(6)</sup>. Let  $\sigma'$  and  $\sigma''$  be defined similarly. Then every integral set  $S$  of  $Q$  has a basis over  $F[z]$  of the form  $1, i, j, \omega$ , where*

$$(1) \quad \omega = ai/\tau' + bj/\sigma' + ij/\sigma'\tau'$$

Presented to the Society, April 8, 1938; received by the editors November 16, 1939.

<sup>(1)</sup> A. A. Albert, *Structure of Algebras*, American Mathematical Society Colloquium Publications, vol. 24, 1939, p. 145.

<sup>(2)</sup> A. A. Albert, *Integral domains of rational generalized quaternion algebras*, Bulletin of the American Mathematical Society, vol. 40 (1934), p. 166.

<sup>(3)</sup> For a summary of results on Hilbert irreducibility fields see A. A. Albert, *Involutorial simple algebras and real Riemann matrices*, Annals of Mathematics, (2), vol. 36 (1935), p. 890.

<sup>(4)</sup> L. E. Dickson, *Algebren und ihre Zahlentheorie*, 1927, p. 155.

<sup>(5)</sup> For analogous results for rational algebras see, in addition to the references already cited, C. G. Latimer, *Arithmetics of generalized quaternion algebras*, American Journal of Mathematics, vol. 48 (1926), pp. 57-66; M. D. Darkow, *Determination of a basis for the integral elements of certain generalized quaternion algebras*, Annals of Mathematics, (2), vol. 28 (1926), pp. 263-270.

<sup>(6)</sup> A polynomial is monic if its leading coefficient is unity.

with  $a, b$  any quantities in  $F[z]$  satisfying

$$(2) \quad \tau b^2 \equiv \tau'^2 \pmod{\sigma'}, \quad \sigma a^2 \equiv \sigma'^2 \pmod{\tau'}.$$

For, by conditions U, C, and R, if  $\xi$  is in an integral set  $S$ , then the traces of  $\xi, i\xi, j\xi$ , and  $ij\xi$  are in  $F[z]$ . Thus  $\xi = x_0 + x_1i/\tau + x_2j/\sigma + x_3ij/\sigma\tau$  with the  $x$ 's all in  $F[z]$ . Now  $N(\xi)$  is in  $F[z]$  if and only if

$$(3) \quad x_1^2\sigma \equiv x_3^2 \pmod{\tau}, \quad x_2^2\tau \equiv x_3^2 \pmod{\sigma}.$$

Since  $\sigma$  is not a quadratic residue of any factor of  $\tau''$  and  $\tau$  is not a quadratic residue of any factor of  $\sigma''$ , we see from the congruences (3) that  $x_1$  is divisible by  $\tau''$ ,  $x_2$  by  $\sigma''$ , and  $x_3$  by  $\sigma''\tau''$ . It follows that every quaternion in an integral set  $S$  lies in a domain  $(1, i, j, \omega)$  over  $F[z]$ , where  $\omega$  is defined in (1).

Let  $r_3/\sigma'\tau'$  be the g.c.d. of all the coefficients of  $ij$  for quantities in  $S$ . Then  $r_3/\sigma'\tau'$  is a linear combination (with multipliers in  $F[z]$ ) of the coefficients of  $ij$  of a finite set of quaternions of  $S$ . Let  $\rho$  be the corresponding linear combination of the same quaternions. Hence  $\rho$  is in  $S$ . Also  $\rho = r_0 + r_1i + r_2j + r_3\omega'$ , where the  $r$ 's are in  $F[z]$ ,  $\omega' = a'i/\tau' + b'j/\sigma' + ij/\sigma'\tau'$ , and  $a', b'$  satisfy (2). If  $\eta$  is also in the integral set  $S$ ,  $\eta = y_0 + y_1i + y_2j + y_3\omega$  and  $y_3$  is divisible by  $r_3$ . Using the fact that  $N(\rho + \eta)$  must be in  $F[z]$ , we find that  $\eta$  is in  $S' = (1, i, j, \omega')$ . It is easily verified that  $S'$  satisfies conditions R, C, U. Inasmuch as it contains the maximal set  $S$ , we have  $S = S'$ . This completes the proof.

If  $\sigma'$  has  $m$  factors and  $\tau'$  has  $n$  factors, then there are  $2^{m+n}$  pairs of incongruent solutions  $a, b$  of (2). The corresponding  $2^{m+n}$  integral sets may be proved distinct by calculating  $N(\omega + \omega')$  for  $\omega \neq \omega'$ .

The monic quantity  $d = \sigma''\tau''$ , although defined by a particular basis, is an invariant of the algebra called the *fundamental number*<sup>(7)</sup> of  $Q$ . It is in fact, except for a factor in  $F$ , the square root of the discriminant of an integral set of  $Q$ . Every integral set is a maximal order of  $Q$  and all maximal orders of  $Q$  have the same discriminant<sup>(8)</sup>. This implies the invariance of the fundamental number  $d$ . We proceed to give a direct proof based upon our definition of  $d$ .

**THEOREM 2.** *The fundamental number  $d = \sigma''\tau''$  of a quaternion algebra  $Q$  is an invariant of the algebra.*

For, let  $Q$  have a normalized basis,  $1, i, j, ij$ , with  $i^2 = \tau, j^2 = \sigma$ . If  $1, i_0, j_0, i_0j_0$  is another normalized basis,  $i_0^2 = \tau_0, j_0^2 = \sigma_0$ , then

$$\begin{aligned} i_0 &= (x_1i + x_2j + x_3ij)/x_4, & (x_1, x_2, x_3) &= 1, \\ j_0 &= (y_1i + y_2j + y_3ij)/y_4, & (y_1, y_2, y_3) &= 1, \end{aligned}$$

where the  $x$ 's and  $y$ 's are in  $F[z]$ .

<sup>(7)</sup> H. Brandt, *Idealtheorie in Quaternionenalgebren*, Mathematische Annalen, vol. 99 (1928), pp. 1-29; C. G. Latimer, *On the fundamental number of a rational generalized quaternion algebra*, Duke Mathematical Journal, vol. 1 (1935), pp. 433-435.

<sup>(8)</sup> M. Deuring, *Algebren*, 1935, p. 88.

Let  $d_1$  be a prime divisor of  $d = \sigma''\tau''$ , the fundamental number corresponding to the basis  $1, i, j, ij$ . We first assume that  $d_1$  divides  $\tau''$ . Then  $d_1$  divides  $x_2y_2$  because  $i_0j_0 + j_0i_0 = 0$ . Now  $d_1$  cannot divide both  $x_2$  and  $x_4$ ; if so, we would have on computing  $\tau_0$

$$x_1^2\tau - x_3^2\sigma\tau \equiv 0 \pmod{d_1},$$

and thus

$$x_1^2 - x_3^2\sigma \equiv 0 \pmod{d_1},$$

an impossibility for  $d_1$  a divisor of  $\tau''$ . Similarly  $d_1$  does not divide both  $y_2$  and  $y_4$ .

Suppose that  $d_1$  divides  $x_2$ . Then  $d_1$  does not divide  $x_4$  and consequently  $d_1$  divides  $\tau_0$ . If  $d_1$  did not divide  $\tau_0''$ , we would have

$$\sigma_0 \equiv c^2 \pmod{d_1}$$

and thus

$$y_1^2\tau + y_2^2\sigma - y_3^2\sigma\tau \equiv c^2y_4^2 \pmod{d_1},$$

$$(4) \quad y_2^2\sigma \equiv c^2y_4^2 \pmod{d_1}.$$

Since  $(\sigma_0, \tau_0) = 1$ , we have  $(\sigma_0, d_1) = 1$  and also  $(c, d_1) = 1$ . Noticing also that  $(\sigma, d_1) = 1$ , we see that congruence (4) implies that  $\sigma$  is a quadratic residue of  $d_1$ , a contradiction to the assumption that  $d_1$  divides  $\tau''$ . We have proved that  $d_1$  divides  $\tau_0''$  and hence that it also divides  $d_0 = \sigma_0''\tau_0''$ .

If  $d_1$  divides  $y_2$ , similar reasoning would show that  $d_1$  divides  $\sigma_0''$ .

A parallel proof is used in case we had assumed  $d_1$  to be a divisor of  $\sigma''$  to demonstrate that  $d_1$  divides either  $\sigma_0''$  or  $\tau_0''$ .

Hence every prime divisor of  $d$  divides  $d_0$  and, of course, conversely. Since  $d$  and  $d_0$  are square-free and monic,  $d = d_0$ .

We shall use this lemma of Albert<sup>(\*)</sup>.

**LEMMA 1.** *If in the generalized quaternion algebra  $Q$  we replace  $\sigma$  by  $(g^2 - \tau h^2)\sigma$  with  $g, h$  in  $F(z)$ , we obtain an equivalent algebra.*

In the remainder of this section  $F$  is specialized to be the field of all real numbers. We apply Lemma 1 to prove

**THEOREM 3.** *Let  $F$  be the field of all real numbers. Then  $Q$  over  $F(z)$  has a basis  $1, i, j, ij$ , with  $i^2 = -1$  and  $j^2 = \sigma$ , where  $\sigma$  has leading coefficient  $\pm 1$ , is a product of distinct linear factors, and is, except for sign, the fundamental number of  $Q$ . There is a single integral set  $S$  and it has a basis  $1, i, j, ij$ . Furthermore there is a one-to-one correspondence between the classes of equivalent quaternion*

(\*) See footnote 2.



algebras (including non-division algebras) over  $F(z)$  and the square-free polynomials  $\sigma$  in  $F[z]$  of leading coefficient  $\pm 1$  containing only linear factors.

By a theorem of Tsen<sup>(10)</sup>, there are no normal division algebras of order greater than 1 over the field of complex numbers with one indeterminate adjoined. Hence  $F((-1)^{1/2})$  splits  $Q$  and we may take  $i^2 = -1$  since  $Q$  contains<sup>(11)</sup> a field equivalent to  $F((-1)^{1/2})$ . Now  $j^2 = \sigma$  and we may take  $\sigma$  square-free and in  $F[z]$ . The leading coefficient may be taken as  $\pm 1$ , since if  $\sigma$  has leading coefficient  $a$  then  $\sigma/(|a|^{1/2})^2$  has the desired property.

If  $r = z^2 + 2bz + c$ , with  $b$  and  $c$  in  $F$ , and is positive definite, the discriminant  $d_0$  of  $r$  is  $4(b^2 - c)$  and is negative. Then  $r$  is a sum of two squares in  $F[z]$ :

$$r = (z + b)^2 + (\frac{1}{2}(-d_0)^{1/2})^2.$$

If  $r$  divides  $\sigma$ , an application of Lemma 1 in reverse when  $\tau = -1$  serves to remove the factor  $r$  from  $\sigma$ . In this way all positive definite prime factors of  $\sigma$  are removed, and we can assume now that  $\sigma$  contains no such factors. The only other irreducible polynomials in  $F[z]$  are linear. Hence  $\sigma$  is a product of linear factors and they are distinct because  $\sigma$  is square-free.

The fact that  $1, i, j, ij$  form a basis of the integral set  $S$  follows immediately from Theorem 1, since  $\tau = -1$  and  $-1$  is never a quadratic residue of a linear function of  $F[z]$ . Hence the fundamental number of  $Q$  is  $\pm\sigma$ .

Let  $\sigma$  have leading coefficient  $\pm 1$  and contain only distinct linear factors. If the norm

$$(5) \quad x_0^2 + x_1^2 - \sigma(x_2^2 + x_3^2)$$

of a quaternion in  $S$  is zero, it must be zero for every value taken in  $F$  by the indeterminate  $z$ . Setting  $z$  in turn equal to each of the roots of  $\sigma$  and using the fact that  $x_0^2 + x_1^2$  is positive definite, we deduce that both  $x_0$  and  $x_1$  are divisible by  $\sigma$ . Dividing (5) by  $\sigma$  and using the same reasoning, we find that  $x_2$  and  $x_3$  are both divisible by  $\sigma$ . Continuing in this way, we find that  $x_0, x_1, x_2, x_3$  are all divisible by every power of  $\sigma$ . This is possible only when  $\sigma$  is in  $F$ , i.e.,  $\sigma = \pm 1$ . But  $\sigma \neq -1$ , for then (5) is positive definite. Consequently when  $Q$  is not a division algebra,  $\sigma = 1$ .

If  $Q$  contains quantities having norms with negative leading coefficient, then using (5) we conclude that  $\sigma$  is monic; otherwise  $-\sigma$  is monic. Hence the sign of  $\sigma$  is determined by the algebra.

We know then that  $\sigma$  is uniquely determined by the algebra since except for sign it is the fundamental number of the algebra.

**3. Integral sets of  $Q$  with characteristic two.** Let the field  $F$  have characteristic two. Then  $Q$  has a basis<sup>(12)</sup>  $1, i, j, ij$  where  $i^2 + i + \alpha = 0$ ,  $j^2 = \gamma$ ,  $ij$

<sup>(10)</sup> C. C. Tsen, *Algebren über Funktionenkörpern*, Göttingen Dissertation, 1934.

<sup>(11)</sup> M. Deuring, *Algebren*, 1935, p. 46.

<sup>(12)</sup> A. A. Albert, *Structure of Algebras*, 1939, p. 145.

$=j(i+1)$ , and  $\alpha$  and  $\gamma$  are in  $F(z)$ . Choose  $m_2 \neq 0$ ,  $m_0$  in  $F(z)$  so that  $\gamma_0 = m_0^2 + \gamma m_2^2$  is in  $F[z]$  and has minimal degree in the set of all quantities of that form. Now  $\gamma_0 \neq 0$  because otherwise the nonzero quaternion  $m_0 + m_2 j$  would be a divisor of zero. Evidently  $\gamma_0$  is square-free. Whenever  $\gamma'_0 = m_0'^2 + \gamma_0 m_2'^2$ , then  $\gamma'_0 = (m_0' + m_0 m_2')^2 + \gamma(m_2 m_2')^2$ ; hence  $\gamma_0$  has minimal degree in the set of all quantities of  $F[z]$  of the form  $m_0'^2 + \gamma_0 m_2'^2$ , where  $m_2' \neq 0$ . The transformation

$$i_1 = i + m_0 i j / \gamma m_2, \quad j_1 = m_0 + m_2 j$$

replaces  $\gamma$  by  $\gamma_0$ .

Let  $\beta_0$  be a nonzero quantity of lowest degree for which the equation  $x j_1 = j_1(x + \beta_0)$  has a solution with  $x$  an integral quaternion. Denote such a solution  $x$  by  $r_0 + r_1 i_1 + r_2 j_1 + r_3 i_1 j_1$ . Necessarily  $r_3 = 0$ . Let  $b'$  be the leading coefficient of  $\beta_0$ . The transformation

$$i_0 = (r_0 + r_1 i_1 + r_2 j_1) / b', \quad j_0 = j_1$$

produces a new basis of  $Q$  of the type described in

**THEOREM 4.** *An algebra  $Q$  of characteristic two has a basis  $1, i, j, ij$  where  $i^2 = \beta i + \alpha$ ,  $j^2 = \gamma$ ,  $ij = j(i + \beta)$ ;  $\alpha, \beta, \gamma$  are in  $F[z]$ ;  $\beta$  is monic and has least degree among all nonzero  $\beta_0$  in  $F[z]$  for which the equation  $xj = j(x + \beta_0)$  has an integral quaternion  $x$  as solution; and  $\gamma$  is a square-free polynomial and has the least degree for all polynomials of the form  $m_0^2 + m_2^2 \gamma$  having  $m_0, m_2$  in  $F(z)$  and  $m_0 \neq 0$ .*

A basis of the type given in Theorem 4 will be called a *normalized basis*.

When  $F$  is perfect we can take  $\gamma = z$ . This is implied by a result of Albert<sup>(15)</sup>. We give here a direct proof. First,  $\gamma$  cannot be in  $F$  for then  $\gamma^{1/2} + j$  would be a divisor of zero. Hence  $\gamma$  has degree  $\geq 1$ . Since  $\gamma$  is in  $F[z]$  and  $F$  is perfect,  $\gamma = c_1^2 + c_2^2 z$  with  $c_1$  and  $c_2$  in  $F[z]$ . Thus  $z = (c_1/c_2)^2 \gamma + (1/c_2)^2$  and has minimal degree. We have proved part of

**THEOREM 5.** *When  $F$  is perfect, then in Theorem 4 we may take  $\gamma = z$  and  $\beta$  monic, square-free, and prime to  $z$ .*

A value of  $\beta$ , because of the minimal degree property, is necessarily square-free. For, if  $\beta = \beta_1 p^2$ , then  $i' = (m_0 + i + m_2 j) / p$  is integral if  $m_0$  and  $m_2$  are chosen in  $F[z]$  to satisfy  $m_0^2 + m_2^2 z = \alpha$ . Furthermore  $i' j = j(i' + \beta / p)$ , and  $\beta / p$  has degree less than that of  $\beta$ . These properties of  $i'$  contradict the assumptions made about  $\beta$ .

In addition,  $\beta$  is not divisible by  $z$ . Otherwise, if we take  $r_0$  in  $F$  to be the square root of the constant term of  $\alpha$ , and  $r_2$  to be the square root of the coefficient of the linear term of  $\beta r_0 + \alpha$ , we have that  $i' = (r_0 + i + r_2 j) / z$  is integral,  $i' j = j(i' + \beta / z)$ , and  $\beta / z$  has degree less than that of  $\beta$ . We have here a contradiction to the defining property of  $\beta$ .

<sup>(15)</sup> A. A. Albert, *p-algebras over a field generated by one indeterminate*, Bulletin of the American Mathematical Society, vol. 43 (1937), p. 735.

**THEOREM 6.** *Let  $Q$  have a normalized basis. Then every integral set  $S$  in  $Q$  has a basis  $1, i, j, \omega = (x_1x_2 + x_1i + x_2j + ij)/m$ , where  $m$  is the largest factor of  $\beta\gamma$  for which there are solutions  $x_1, x_2$  of*

$$x_1^2 \equiv \gamma, \quad x_2^2 \equiv \beta x_2 + \alpha \pmod{m}.$$

Furthermore  $m$  is square-free.

Let  $\xi$  be in  $S$ . By properties R, C, and U, the traces of the quaternions  $\xi, i\xi, j\xi, ij\xi$  are in  $F[z]$ . Hence  $\xi = [x_0 + x_1i + (x_2 + x_3i)j/\gamma]/\beta$  with the  $x$ 's in  $F[z]$ .

Since the denominators of integral quantities divide  $\beta\gamma$ , an integral set  $S$  must have a basis. This basis can be taken in the form

$$\begin{aligned} \omega_1 &= e_0/\beta\gamma, & \omega_2 &= (f_0 + f_1i)/\beta\gamma, \\ \omega_3 &= (g_0 + g_1i + g_2j)/\beta\gamma, & \omega_4 &= (h_0 + h_1i + h_2j + h_3ij)/\beta\gamma, \end{aligned}$$

with the  $e$ 's,  $f$ 's,  $g$ 's, and  $h$ 's in  $F[z]$ . We may assume, since  $1, i, j, ij$  are all in  $S$ , that  $e_0, f_1, g_2, h_3$  either equal  $\beta\gamma$  or else have degree less than that of  $\beta\gamma$  and the remaining  $f_0, g_1$ 's, and  $h$ 's have degrees less than  $D(\beta\gamma)$  (the degree of a polynomial  $a$  is designated by  $D(a)$ ).

Obviously,  $\omega_1$  is not integral unless  $e_0 = \beta\gamma$ ;  $\omega_1 = 1$ .

If  $D(f_1) < D(\beta\gamma)$ , then  $D(T(\omega_2)) < D(\beta)$  while  $\omega_2j = j(\omega_2 + T(\omega_2))$ . This contradicts the choice of  $\beta$ ; hence  $D(f_1) = D(\beta\gamma)$  and in fact  $f_1 = \beta\gamma$ . Since  $\omega_2 - i$  is in  $S$ ,  $f_0/\beta\gamma$  is in  $F[z]$ ; hence  $f_0 = 0$ , and  $\omega_2 = i$ .

From  $D(g_1) < D(\beta\gamma)$ , it follows that  $D(T(\omega_3)) < D(\beta)$  and  $\omega_3j = j(\omega_3 + T(\omega_3))$ , a contradiction to the choice of  $\beta$  unless  $g_1 = 0$ . If  $D(g_2) < D(\beta\gamma)$ , then  $\omega_3$  has its norm  $(g_0/\beta\gamma)^2 + (g_2/\beta\gamma)^2\gamma$  in  $F[z]$  and of degree less than that of  $\gamma$ , a contradiction to the choice of  $\gamma$ . Thus  $g_2 = \beta\gamma$ , and since  $\omega_3 - j$  is in  $S$ ,  $g_0 = 0$ , so that  $\omega_3 = j$ .

Since  $ij$  is in  $S$ , necessarily  $h_3$  divides  $\beta\gamma$ ;  $\beta\gamma = h_3m'$  with  $m'$  in  $F[z]$ . Now  $\omega_4m' - ij = h_0/h_3 + h_1i/h_3 + h_2j/h_3$  is in  $S$ . Thus  $h_0, h_1, h_2$  are all divisible by  $h_3$  and  $\omega_4 = (d_0 + d_1i + d_2j + ij)/m'$  with the  $d$ 's in  $F[z]$ . In  $S$  must be  $i\omega_4$  and  $\omega_4j$ . This is possible if and only if

$$(6) \quad d_1^2 \equiv \gamma, \quad d_2^2 \equiv d_2\beta + \alpha, \quad d_0 \equiv d_1d_2 \pmod{m'}.$$

Now  $m'$  has no square factors. Otherwise, if  $p^2$  were a divisor of  $m$ ,  $p^2$  would divide  $\beta\gamma$ . If  $p$  were a divisor of  $\gamma$ , then because  $d_1^2 \equiv \gamma \pmod{p^2}$ ,  $p$  would divide  $d_1$  and  $p^2$  divide the square-free  $\gamma$ . Hence  $p$  would not divide  $\gamma$ , so that  $p^2$  would be a divisor of  $\beta$ . But then  $i_1 = (d_2 + i)/p$  would be integral,  $i_1j = j(i_1 + \beta/p)$ , and  $\beta/p$  have smaller degree than  $\beta$ . This is impossible from our choice of  $\beta$ .

Our next step is to give a construction of  $\omega_4$ . Let  $m$  be the product of all prime powers  $p_n^a$  dividing  $\beta\gamma$  for which there exist solutions of

$$(7) \quad x_{1n}^2 \equiv \gamma, \quad x_{2n}^2 \equiv x_{2n}\beta + \alpha \quad (p_n^{e_n}).$$

By means of a discussion similar to that for  $m'$  we can show that  $m$  is also square-free, i.e.,  $e_n = 1$ . Using the Chinese remainder theorem we can find a unique solution  $x_1, x_2$  modulo  $m$  of the congruences (7) common to all  $p_n$ . Therefore

$$x_1^2 \equiv \gamma, \quad x_2^2 \equiv x_2\beta + \alpha \quad (m).$$

The quantity  $\omega = (x_1x_2 + x_1i + x_2j + ij)/m$  is integral. Its trace is  $\beta x_1/m$ . This is in  $F[z]$  since any factor of  $m$  dividing  $\gamma$  divides  $x_1$  because of (7) and the remaining factors of  $m$  divide  $\beta$ . Furthermore

$$\begin{aligned} m^2 N(\omega) &= x_1^2 x_2^2 + x_1^2 x_2 \beta + x_1^2 \alpha + \gamma(x_2^2 + x_2 \beta + \alpha) \\ &= (x_1^2 + \gamma)(x_2^2 + x_2 \beta + \alpha) \equiv 0 \quad (m^2); \end{aligned}$$

thus  $N(\omega)$  is in  $F[z]$ .

The quantity  $\omega$  with  $1, i, j$  forms a basis for an integral set  $S'$ . The conditions C, R, and U are easily verified to be satisfied. To show that maximality is true only for such a set  $S'$ , we need only show that every integral set  $S$  is necessarily contained in such a set; in fact, only that  $\omega_4$  is in some  $S'$ .

Since (6) holds for  $m'$ , it is true of every prime factor of  $m'$ . Also  $m'$  divides  $\beta\gamma$ . From the definition of  $m$ , every prime factor of  $m'$  divides  $m$ . Thus  $m'$  divides  $m$ ;  $m = m'm''$ . We can find a solution  $x_1, x_2$  of (7) for which  $x_1 \equiv d_1, x_2 \equiv d_2 \pmod{p_n}$  whenever  $p_n$  is a divisor of  $m'$ . Consequently  $\omega_4$  is in  $(1, i, j, \omega m'')$  which is in  $S$ . We have proved our theorem.

Another form for the basis of  $Q$  of characteristic two<sup>(14)</sup> is  $1, u_1, u_2, u_1u_2$ , where

$$u_1^2 = \tau, \quad u_2^2 = \sigma, \quad u_1u_2 + u_2u_1 = \rho \quad (\rho, \sigma, \tau \text{ in } F(z)).$$

Such a basis can be obtained by taking  $u_1 = j, u_2 = ij$ . A basis of  $Q$  of this form can be found which is normalized to have  $\rho, \sigma, \tau$  in  $F[z]$ ,  $u_1$  a quantity with norm of lowest degree in the set of all inseparable integral quantities over  $F(z)$ , and  $u_2$  an integral quantity linearly independent of 1 and  $u_1$ , inseparable over  $F(z)$ , and having for  $\rho$  a value in  $F[z]$  of lowest degree. Using much the same reasoning as before we can prove

**THEOREM 7.** *An integral set  $S$  with respect to a basis  $1, u_1, u_2, u_1u_2$  normalized as above has a basis  $1, u_1, u_2, \omega$ , where*

$$\omega = (y_1 + y_1u_1 + y_2u_2 + y_3u_1u_2)/\rho.$$

<sup>(14)</sup> N. Jacobson, *p-algebras of exponent p*, Bulletin of the American Mathematical Society, vol. 43 (1937), pp. 667-670.

Here  $y_3$  is determined as one of the quantities of lowest degree for which there exists a solution of

$$y_0^2 + y_1^2\tau + y_2^2\sigma + y_3^2\sigma\tau + \rho(y_0y_3 + y_1y_2) = 0 \quad (\rho^2)$$

with the  $y$ 's in  $F[z]$ .

4. **Factorization when  $S$  is a principal ideal ring.** Theorems 8 and 9 give sufficient conditions for an integral set  $S$  to possess a weakened form of a Euclidean algorithm. This form of the algorithm, however, is equivalent to the algorithm itself for quaternion algebras.

**THEOREM 8.** *Let  $Q$  of characteristic not two have a normalized basis with  $\sigma, \tau$  having degrees not greater than 1, and if both have degree 1, then one of them being a quadratic residue of the other. Then if  $\theta$  is in an integral set  $S$  of  $Q$ , and  $m$  is a nonzero polynomial in  $F[z]$ , there exists a quaternion  $\kappa$  in  $S$  such that  $D(N(\theta - \kappa m)) < D(N(m))$ .*

A proof of this theorem is easily effected when an explicit basis of  $S$  is known.

If  $\sigma$  and  $\tau$  are both in  $F$ ,  $S = (1, i, j, ij)$ .

Suppose  $\sigma$  is linear and  $\tau$  is in  $F$ . Were  $\tau$  a quadratic residue of  $\sigma$ , we would have  $\tau = a^2$  and  $Q$  would not be a division algebra. Hence  $\tau$  is not a quadratic residue of  $\sigma$  and  $S = (1, i, j, ij)$ . The case  $\sigma$  in  $F$  and  $\tau$  linear is treated similarly.

Suppose that both  $\sigma$  and  $\tau$  are linear. If  $(\sigma|\tau) = (\tau|\sigma) = -1$  (this case is excluded in the theorem),  $S = (1, i, j, ij)$ . If however  $(\sigma|\tau) = -(\tau|\sigma) = 1$ , then  $\sigma \equiv a^2(\tau)$ , with  $a$  in  $F$ . Hence  $S = (1, i, j, \omega)$ , where  $\omega$  is one of  $i(a \pm j)/\tau$ . The case  $(\tau|\sigma) = -(\sigma|\tau) = 1$  is handled similarly. Finally if  $(\sigma|\tau) = (\tau|\sigma) = 1$ , then  $\sigma \equiv a^2(\tau)$ ,  $\tau \equiv b^2(\sigma)$ , with  $a$  and  $b$  in  $F$ . Thus  $\sigma = a^2 + k\tau$ ,  $b^2 = -a^2/k$ , and  $i/a\tau + j/b\sigma + ij/\sigma\tau$  has norm 0;  $Q$  is total matric.

If in Theorem 8 we write  $\theta = g_0 + g_1i + g_2j + g_3\omega$ , the quaternion  $\kappa = q_0 + q_1i + q_2j + q_3\omega$  is found by choosing the polynomials  $q_k$  to satisfy  $D(g_k - q_k m) < D(m)$ ; i.e., the  $q_k$  are the quotients on dividing the  $g_k$  by  $m$ .

**THEOREM 9.** *Let  $Q$  have characteristic two, with  $\gamma$  linear and  $\alpha$  and  $\beta$  in  $F$ . If  $\theta$  is in the integral set  $S = (1, i, j, ij)$  of  $Q$ , and  $m$  is in  $F[z]$ , then there exists a quaternion  $\kappa$  in  $S$  such that  $D(N(\theta - \kappa m)) < D(m^2)$ .*

That  $S$  has a basis  $1, i, j, ij$  follows from the discussion in §3 and the fact that  $x^2 + \beta x + \alpha = N(x+i)$  is irreducible in  $F$  when  $Q$  is a division algebra. The quaternion  $\kappa$  is determined as in the proof of Theorem 8.

When  $Q$  has characteristic two and  $F$  is perfect, we can take  $\gamma = z$  by Theorem 5. If in addition  $\beta$  is in  $F$ , then we can assume  $\beta = 1$ . We can also have  $\alpha$  in  $F$ . For, since  $F$  is perfect,  $\alpha$  has the form  $a_1^2 + a_2^2z$ . The degree of  $\alpha$  is reduced to zero by repeated application of the transformation

$$i' = a_1 + i + a_2j, \quad j' = j.$$



We then have a basis for this  $Q$  satisfying the hypothesis of Theorem 9.

Theorems 8 and 9 imply the existence of a Euclidean algorithm for the integral sets involved<sup>(15)</sup>. The presence of such a process assures us that  $S$  is a principal ideal ring.

Whenever  $S$  is a principal ideal ring, the following decomposition theorem is true. A proof can be made using a procedure developed for rational algebras<sup>(16)</sup>.

**THEOREM 10.** *Let  $S$  be a principal ideal ring. Let  $\theta$  be a quaternion in  $S$  not divisible by a polynomial in  $F[z]$ . If  $N(\theta) = p_1 p_2 \cdots p_n$ , where the  $p_k$  are irreducible polynomials, then  $\theta = \pi_1 \pi_2 \cdots \pi_n$  where  $N(\pi_k) = p_k$  and  $\pi_1$  is unique except for multiplication by units of  $S$  on the right,  $\pi_2, \dots, \pi_{n-1}$  are unique but for multiplication by units on the right or left, and  $\pi_n$  is unique except for left unit factors.*

**5. Prime quaternions in  $S$ .** In this section we seek to determine when a quaternion is prime in  $S$ . In particular we want to know when a prime in  $F[z]$  is prime in  $S$ . All ideals considered are left ideals.

**THEOREM 11.** *Let  $F$  have characteristic not two. Then a necessary and sufficient condition that the principal ideal  $(p)$  defined by a prime  $p$  of  $F[z]$  not dividing the fundamental number  $d$  of  $Q$  be divisorless in  $S$  is that there exist no solution in  $F[z]$  of the congruence*

$$(8) \quad 1 - x_1^2 \tau - x_2^2 \sigma + x_3 \sigma \tau \equiv 0 \pmod{p}.$$

If (8) holds, let

$$(9) \quad \xi = 1 + x_1 i + x_2 j + x_3 ij,$$

and let  $P$  be the left ideal  $(\xi, p)$ ;  $P$  is a proper divisor of  $(p)$ . Also  $P \neq (1)$ . For otherwise  $1 = \alpha \xi + \beta p$  with  $\alpha, \beta$  in  $S$ ,  $\bar{\xi} = \alpha(\bar{\xi}\bar{\xi}) + \beta \bar{\xi} p \equiv 0 \pmod{p}$ , an impossibility. Hence  $(p)$  is not a divisorless ideal.

Conversely, suppose there is a left ideal  $P \neq (1)$  which properly divides  $(p)$ ; i.e.,  $P$  contains a quaternion  $\xi$  not divisible by  $p$ . Necessarily  $N(\xi)$  is divisible by  $p$ . If  $p$  does not divide  $\sigma'\tau'$ , by multiplying  $\xi$  by  $i, j$ , or  $ij$  if necessary, we can obtain an element  $\xi_0$  whose coefficient  $x_0$  of 1 is not congruent to 0  $\pmod{p}$ . We can find a solution  $m$  in  $F[z]$  of  $m\sigma'\tau'x_0 \equiv 1 \pmod{p}$ ,  $m\sigma'\tau'x_0 = 1 + rp$ . Then  $\xi_0 m\sigma'\tau' - rp = 1 + y_1 i + y_2 j + y_3 ij$  has norm congruent to 0  $\pmod{p}$ , and the  $y_k$  are in  $F[z]$ . Hence congruence (8) has a solution. If  $p$  divides  $\sigma'\tau'$ , then from the property of such a prime factor we know that (8) has a solution.

If the ideals of  $S$  are all principal, then  $kp = \pi_1 \bar{\pi}_1$ , where  $k$  is in  $F$  and

<sup>(15)</sup> H. Rauter, *Quaternionenalgebren mit Komponenten aus einem Körper von Primzahlcharakteristik*, Mathematische Zeitschrift, vol. 29 (1929), pp. 234-263.

<sup>(16)</sup> C. G. Latimer, *On ideals in generalized quaternion algebras and Hermitian forms*, these Transactions, vol. 38 (1935), pp. 443-444.

$(\pi_1) = (\xi, p)$  with  $\xi$  defined in (9). Also  $\pi_1$  is a divisorless quaternion of  $S$  because its norm is a prime in  $F[z]$ .

It is known<sup>(17)</sup> that only the prime divisors of the fundamental number  $d$  are ramified in  $S$ .

**THEOREM 12.** *Every prime divisor  $p$  of the fundamental number of  $Q$  of characteristic not two generates an ideal in  $S$  which is the square of a two-sided prime ideal  $R$ .*

A proof of this theorem can be made by following the steps in the demonstration of the analogous theorem for rational quaternion algebras by A. Spaltenstein<sup>(18)</sup>. Let  $S_p$  denote the difference algebra  $S - (p)$  where  $p$  is in  $F[z]$ . If  $p$  is a prime dividing the fundamental number  $d$  of  $Q$ , then  $S_p$  contains a unique nonzero idempotent element. Using this fact we can prove that the radical  $R_p$  of  $S_p$  has exponent two and is the only maximal proper two-sided ideal in  $S_p$ . The ideal  $R$  in Theorem 12 is the set of quantities of  $S$  in the residue classes comprising  $R_p$ .

**THEOREM<sup>(19)</sup> 13.** *Let  $Q$  be over  $F(z)$ , where  $F$  is a finite field. Then no prime of  $F[z]$  generates a prime ideal in  $S$ .*

First, let  $F$  have characteristic not two. If a quantity  $p$  of  $F[z]$  generates a prime ideal in  $Q$ , it does not divide the discriminant of  $S$ , as a result of Theorem 12. Then  $S_p$  is semisimple. Also since  $(p)$  is prime in  $S$ ,  $S_p$  contains no divisors of zero; hence  $S_p$  is a division algebra and because it is also finite, it is a field. Thus  $-ij = ji \equiv ij (p)$ ; whence  $2 \equiv 0 (p)$ , an impossibility. If  $F$  has characteristic 2, we may take  $\gamma = z$ . Every quantity in  $F[z]$  has the form  $f(z^2) + g(z^2) \cdot z = f(z)^2 + g(z)^2 \cdot z$  and is therefore the norm of  $f(z) + g(z) \cdot j$ . This completes the proof of our theorem.

By a result of Eichler<sup>(20)</sup>, every ideal in  $S$  is principal when  $F$  is a finite field. This fact, together with Theorem 13, gives

**THEOREM 14.** *When  $F$  is finite, every polynomial in  $F[z]$ , except for a factor in  $F$ , is the norm of a quaternion in  $S$ .*

As a particular instance we have that every polynomial in  $F[z]$  is expressible in the form  $x_0^2 - fx_1^2 \pm (z - g)(x_2^2 - fx_3^2)$  where  $f$  is a non-square fixed quantity in  $F$ ,  $g$  is fixed in  $F$ , and the  $x$ 's take values in  $F[z]$ .

Combining the results of Theorems 11 and 14 for  $F$  finite and of characteristic not two, we see that congruence (8) always has a solution if  $p$  does

<sup>(17)</sup> M. Deuring, *Algebren*, 1935, p. 84.

<sup>(18)</sup> A. Spaltenstein, *Struktur und Zahlentheorie einer Klasse von Algebren*, Zurich Dissertation, 1934, p. 24.

<sup>(19)</sup> For the rational analogue see A. Speiser, "Idealtheorie in rationalen Algebren," in L. E. Dickson, *Algebren und ihre Zahlentheorie*, 1927, p. 302.

<sup>(20)</sup> M. Eichler, *Über die Idealklassenzahl hyperkomplexer Systeme*, Mathematische Zeitschrift, vol. 43 (1938), pp. 481-494.

not divide the fundamental number. It can be shown,\* however, that if  $(p, \sigma) = 1$ , there is a solution of

$$x^2 - \sigma y^2 - \tau \equiv 0 \pmod{p},$$

a more inclusive fact.

For the remainder of this section we restrict  $F$  to be the field of all real numbers; hence we can take  $\tau = -1$ . The primes in  $F[z]$  are either linear or definite quadratic.

If  $p$  is positive definite,  $p = z^2 + 2rz + s$ , and the ideal generated by  $(r^2 - s)^{1/2} + (z + r)i$  properly contains  $(p)$ ; hence  $(p)$  is not a divisorless ideal of  $S$ .

If  $p$  is linear,  $p = z - a$ , and if there is a solution of the congruence (8), then evaluating the left member at  $z = a$ , we get the necessary condition that the polynomial  $\sigma = \sigma(z)$  must have a positive value for  $z = a$ . Conversely, if  $\sigma(a) > 0$  and  $p = z - a$ , a solution of (8) exists; e.g.,  $x_1 = 0 = x_2$ ,  $x_2 = (1/\sigma(a))^{1/2}$ . We have proved

**THEOREM 15.** *Let  $Q$  be a generalized quaternion algebra over the field  $F(z)$ , where  $F$  is the field of all real numbers. A quantity  $p(z)$  of  $F[z]$  generates a divisorless ideal in the integral set  $S$  of  $Q$  with respect to a normalized basis if and only if  $p(z)$  is linear and the root of  $p(z) = 0$  gives  $\sigma$  a negative value.*

As a result of Theorem 8 we know that  $S$  is a principal ideal ring if  $\sigma$  is linear. This and the fact that the product of two norms is a norm give

**COROLLARY.** *If and only if all the monic linear factors of a square-free polynomial  $f$  in  $F[z]$  have their constant terms not less than  $c$ , then*

$$f = \pm [x_0^2 + x_1^2 - (x_2^2 + x_3^2)(z - c)] \quad (x_k \text{ in } F[z])$$

*If and only if all the constant terms are not greater than  $c$ ,*

$$f = \pm [x_0^2 + x_1^2 + (x_2^2 + x_3^2)(z - c)] \quad (x_k \text{ in } F[z]).$$

If  $\sigma = -1 = \tau$ , then the left member of (8) is always positive for any value of  $z$ . It is never divisible by a linear polynomial. Using this fact and the result that  $S$  is a principal ideal ring, we obtain

**THEOREM 16.** *Let  $F$  be the field of all real numbers, and  $\sigma = -1 = \tau$ . Then every linear polynomial in  $F[z]$  is prime in  $S$ , and every irreducible quadratic polynomial is, except for sign, the norm of a quaternion in  $S$ .*

When  $F$  is the rational number field, there are some positive definite polynomials<sup>(21)</sup>, e.g.,  $z^2 + 7$ , which are prime in  $S$  with  $i^2 = -1 = j^2$ .

<sup>(21)</sup> E. Landau, *Über die Zerlegung definiter Funktionen in Quadrate*, Archiv der Mathematik und Physik, (3), vol. 7 (1904), pp. 271-277.

6. **Equivalence of Hermitian forms and left ideals**<sup>(22)</sup>. Denote by  $G$  the integral domain  $F[z, i]$ ; if the basis of  $Q$  is normalized,  $G$  is the set of all integral elements in the quadratic extension  $F(z, i)$  of the field  $F(z)$ . Let  $W$  designate the set of all quaternions  $\kappa = q_0 + q_1i + q_2j + q_3ij$  with components  $q_i$  in  $F[z]$ ;  $W$  has a basis  $1, j$  over  $G$ . Thus  $\kappa = p_1 + p_2j$  with  $p_1, p_2$  in  $G$  and

$$N(\kappa) = \begin{vmatrix} p_1 & p_2 \\ \eta \bar{p}_2 & \bar{p}_1 \end{vmatrix},$$

where  $\eta = \gamma$  or  $\sigma$  according as  $F$  has or has not characteristic two. The conjugate of a quantity  $w$  of  $G$  is written  $\bar{w}$ .

A left ideal  $L$  of  $W$  is called *regular* if it has a basis (called a *regular basis*)  $\omega_1, \omega_2$  over  $G$  where  $\omega_m = g_{m1} + g_{m2}j$  ( $m = 1, 2$ ) with the  $g_{mn}$  in  $G$  and the determinant  $|g_{mn}|$  in  $F[z]$  and monic. The value of the determinant  $|g_{mn}|$  is independent of the basis  $\omega_1, \omega_2$  and is the *norm*  $N(L)$  of  $L$ . A left ideal  $L$  is said to be *equivalent* to a left ideal  $L'$  if there exist quantities  $\rho, \rho'$  in  $W$  for which  $L\rho = L'\rho'$  and  $N(\rho\rho')$  is monic.

A form

$$(10) \quad f(x_1, x_2) = ax_1\bar{x}_1 + b\bar{x}_1x_2 + \bar{b}x_1\bar{x}_2 + cx_2\bar{x}_2$$

with  $a, c$  in  $F[z]$  and  $b$  in  $G$  is called a Hermitian form of  $G$  and its determinant is defined to be  $b\bar{b} - ac$ . We suppose that the  $x$ 's run over elements of  $G$ . If another Hermitian form  $f'(y_1, y_2)$  can be obtained from  $f(x_1, x_2)$  by a linear homogeneous transformation of determinant unity with coefficients in  $G$ , then  $f$  and  $f'$  are called *equivalent*.

Let  $L$  be a regular ideal with the regular basis  $\omega_m = g_{m1} + g_{m2}j$  ( $m = 1, 2$ ). Since  $j\omega_1, j\omega_2$  are in  $L$ ,

$$(11) \quad j\omega_m = b_{m1}\omega_1 + b_{m2}\omega_2 \quad (m = 1, 2),$$

where the  $b$ 's are in  $G$ . If we designate the general element of  $L$  by  $\xi$ ,

$$\begin{aligned} \xi &= x_1\omega_1 + x_2\omega_2 = (g_{11}x_1 + g_{21}x_2) + (g_{12}x_1 + g_{22}x_2)j, \\ j\xi &= c_1\omega_1 + c_2\omega_2 = (g_{11}c_1 + g_{21}c_2) + (g_{12}c_1 + g_{22}c_2)j, \end{aligned}$$

where  $c_n = b_{1n}\bar{x}_1 + b_{2n}\bar{x}_2$  ( $n = 1, 2$ ), and  $x_1, x_2$  are in  $G$ . Then

$$N(\xi) = \begin{vmatrix} x_1 & x_2 \\ c_1 & c_2 \end{vmatrix} \cdot \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = N(L) \cdot f(x_1, x_2),$$

<sup>(22)</sup> For the rational analogue see C. G. Latimer, *On ideals in generalized quaternion algebras and Hermitian forms*, these Transactions, vol. 38 (1935), pp. 436-446; C. G. Latimer, *On ideals in a quaternion algebra and the representation of integers by Hermitian forms*, these Transactions, vol. 40 (1936), pp. 439-449; C. G. Latimer, *On the class number of a quaternion algebra with a negative fundamental number*, these Transactions, vol. 40 (1936), pp. 318-323; J. D. H. Teller, *A class of quaternion algebras*, Duke Mathematical Journal, vol. 2 (1936), pp. 280-286.

where

$$(12) \quad f(x_1, x_2) = \begin{vmatrix} x_1 & x_2 \\ c_1 & c_2 \end{vmatrix} = b_{12}x_1\bar{x}_1 - b_{11}x_1x_2 + b_{22}x_1\bar{x}_2 - b_{21}x_2\bar{x}_2.$$

Since  $N(\xi)$  and  $N(L)$  are in  $F[z]$ ,  $f(x_1, x_2)$  is in  $F(z)$  for  $x_1, x_2$  in  $G$ . Since  $f$  is a polynomial in  $G$ , it takes values in  $G$ . Hence  $f(x_1, x_2)$  takes values in  $F[z]$  for  $x_1, x_2$  in  $G$ , and  $f$  is consequently Hermitian. We say that  $f$  corresponds to the regular basis  $\omega_1, \omega_2$ .

The relation between classes of ideals and of forms is described in

**THEOREM<sup>(22)</sup> 17.** *There is a one-to-one correspondence between the classes of regular ideals of  $W$  over  $G$  and the classes of Hermitian forms with determinant  $\eta$  representing a monic quantity in  $F[z]$ .*

We next prove

**LEMMA 2.** *An ideal  $L$  of  $W$  is principal if and only if it is regular and any Hermitian form  $f(x_1, x_2)$  corresponding to it represents a nonzero quantity in  $F$ .*

Let  $f(x_1, x_2)$  correspond to a regular ideal  $L = (\omega_1, \omega_2)$  and  $f(r_1, r_2) = a_0$  in  $F$ . Then  $a_0 = b_{12}r_1\bar{r}_1 - b_{11}r_1r_2 + b_{22}r_1\bar{r}_2 - b_{21}r_2\bar{r}_2$ , where the  $b_{mn}$  are defined by (11). If  $\rho = r_1\omega_1 + r_2\omega_2$ , then  $N(\rho) = a_0N(L)$ . The transformation

$$\begin{aligned} \rho &= r_1\omega_1 + r_2\omega_2, \\ \rho' &= (\bar{r}_1b_{11} + \bar{r}_2b_{21})\omega_1/a_0 + (\bar{r}_1b_{12} + \bar{r}_2b_{22})\omega_2/a_0 \end{aligned}$$

has determinant 1, so that  $\rho, \rho'$  is a regular basis of  $L$ . But  $\rho' = j\rho/a_0$ . Hence  $L = (\rho)$ .

Conversely, if  $L = (\rho)$ ,  $L$  has the regular basis  $(\rho, j\rho/r_0)$ , where  $r_0$  is the leading coefficient of  $N(\rho)$ . To this basis corresponds  $f(x_1, x_2) = r_0x_1\bar{x}_1 - (\rho/r_0)x_2\bar{x}_2$ , which represents  $r_0 = f(1, 0)$  in  $F$ .

Noticing that in the last paragraph  $r_0$  determines  $f$ , we have the

**COROLLARY.** *The number of classes of principal ideals is equal to the index of the group of all quantities of  $F$  which are leading coefficients of norms of unit quaternions in  $W$ , in the group of all quantities of  $F$  which are leading coefficients of norms of quaternions in  $W$ .*

We also have the

**COROLLARY.** *If  $W$  is a principal ideal ring, every ideal is regular.*

We next state

**THEOREM 18.** *A necessary condition that every ideal in  $W$  be principal is that  $W$  be an integral set  $S$ .*

<sup>(22)</sup> C. G. Latimer, *On ideals in generalized quaternion algebras and Hermitian forms*, these Transactions, vol. 38 (1935), p. 442.



Let  $\zeta$  equal  $\beta\gamma$  or  $\sigma\tau$  according as  $F$  has or has not characteristic two, respectively. Suppose that every ideal in  $W$  is principal. Now  $S\zeta$  is in  $W$ , and, since  $W(S\zeta) \leq S(S\zeta) = S\zeta$ , we conclude that  $S\zeta$  is an ideal in  $W$ . Therefore  $S\zeta = W\omega$ , with  $\omega$  in  $W$ . Since 1 is in both  $S$  and  $W$ , we have  $\zeta = \mu\omega$  and  $\nu\zeta = \omega$  with  $\mu$  in  $W$  and  $\nu$  in  $S$ . Then  $\nu\mu\omega = \omega$ , so that  $\nu\mu = 1$ , and  $\nu, \mu$  are units in  $W$ . Next,  $S\zeta = W\omega = (W\mu)\omega = W\zeta$ . Finally  $W = S$ .

The conditions of Theorem 18 and the second corollary of Theorem 17 are by no means sufficient as results at the end of this section show.

Every Hermitian form  $f'$  of determinant  $\eta$  is equivalent to a form  $f = ax_1x_1 + bx_1x_2 + \bar{b}x_2x_1 + cx_2x_2$  with  $D(b_0) < D(a)$ ,  $D(b_1) < D(a) \leq D(c)$ , where  $b = b_0 + b_1i$ . This result is obtained by successive applications of the two transformations  $x'_1 = x_1 + bx_2$ ,  $x'_2 = x_2$ ; and  $x'_1 = x_2$ ,  $x'_2 = -x_1$ .

We assume in the next two paragraphs that  $D(\alpha)$  and  $D(\beta)$  are not greater than 1, or  $D(\tau) \leq 1$ , according as  $F$  has or has not characteristic two. Then  $D(a) + D(c) = D(\eta)$ .

If also  $D(\eta) \leq 1$ , then  $D(a) = 0$ . We see that  $f$  represents a quantity in  $F$ , and if  $f$  corresponds to  $L$ ,  $L$  is principal. Every regular ideal in  $W$  over  $G$  is principal.

But if  $D(\eta) = 2$ , then  $D(a) \leq 1$ , and  $b_0$  and  $b_1$  are in  $F$ . If  $D(a) = 1$ ,  $\eta$  is monic, and  $a_0, c_0$  are the leading coefficients of  $a, c$ , respectively, then  $a_0c_0 = -1$ . Hence  $f(1, a_0)$  is in  $F$  and consequently  $f$  corresponds to a class of principal ideals. This is also true of  $f$  if  $D(a) = 0$ . We conclude that all regular ideals of  $W$  over  $G$  are principal when  $\eta$  is monic and quadratic.

Now let  $F$  be a field in which not every quantity is a square. Examples of such fields are subfields of real numbers. Also if  $F$  is finite of characteristic not two, then  $F$  contains non-square quantities. For, corresponding to each square  $a^2$ , there are two distinct elements  $a, -a$  in the field—the set of squares does not exhaust the field.

**THEOREM 19.** *Let  $F$  be a field of characteristic not two in which not every quantity is a square. Let  $\sigma$  in  $F[z]$  be of odd degree and reducible, and  $\tau$  in  $F[z]$  of even degree with leading coefficient not a square. Then the regular ideals of  $W$  are not all principal.*

Let  $\sigma = \sigma_1\sigma_2$ , where  $\sigma_1, \sigma_2$  are in  $F[z]$  and not in  $F$ . Then the Hermitian form

$$(13) \quad f = \sigma_1x_1x_1 - \sigma_2x_2x_2 = \sigma_1(y_0^2 - \tau y_1^2) - \sigma_2(y_2^2 - \tau y_3^2),$$

where  $x_1 = y_0 + y_1i$ ,  $x_2 = y_2 + y_3i$ , with the  $y$ 's in  $F[z]$ , and has determinant  $\sigma$  and  $f$  does not represent a quantity in  $F$ . For,  $y_0^2 - \tau y_1^2$  and  $y_2^2 - \tau y_3^2$  both have even degrees and one of  $\sigma_1, \sigma_2$  has even degree and the other odd degree. Thus  $f(x_1, x_2)$  for  $(x_1, x_2) \neq (0, 0)$  has degree at least that of one of  $\sigma_1, \sigma_2$ .

**THEOREM 20.** *Let  $F$  be the field of all real numbers. The integral set  $S$  with*



respect to the normalized basis of Theorem 3 is equal to  $W$  and is a principal ideal ring if and only if  $\sigma$  has degree not greater than 1, or degree 2 with positive leading coefficient.

The fact that  $S$  is a principal ideal ring when  $D(\sigma) \leq 1$  is a consequence of Theorem 8.

When  $\sigma$  has odd degree greater than 1, Theorem 19 states that  $W$  is not a principal ideal ring.

If  $D(\sigma) = n \geq 4$  is even and  $\sigma$  has leading coefficient  $+1$ , then

$$f = \sigma_1 \sigma_3 \cdots \sigma_{n-1} x_1 \bar{x}_1 - \sigma_2 \sigma_4 \cdots \sigma_n x_2 \bar{x}_2,$$

where the  $\sigma_i = z - a_i$  ( $a_i < a_{i-1}$ ) are the linear factors of  $\sigma$ , cannot represent a nonzero quantity in  $F$ . For, if we set  $z = a_1$ ,  $f$  is always negative or zero; and for  $z = a_n$ ,  $f$  is always positive or zero.

If  $n \geq 2$  is even and  $\sigma$  has leading coefficient  $-1$ , and if  $\sigma_1, \sigma_2$  are two non-constant factors of  $\sigma = \sigma_1 \sigma_2$ , then (13) cannot represent a quantity in  $F$  because the two terms  $\sigma_1 x_1 \bar{x}_1$  and  $-\sigma_2 x_2 \bar{x}_2$  have leading coefficients of the same sign—there can be no reduction in degree by adding values of these two terms.

We have already shown that when the degree of  $\sigma$  is 2 and  $\sigma$  is monic that the regular ideals of  $W$  are principal. It remains to prove that every ideal in  $W$  is regular. We can find a basis  $a, b + jc$  of an ideal  $L$  with  $a, b, c$  in  $G$ , since  $G$  is a principal ideal ring. Since  $ja$  and  $j(b + jc)$  are in  $L$ ,  $a = a_1 c$ ,  $b = b_1 c$ . The ideal  $L_1 = (a_1, b_1 + j)$  is equivalent to  $L$  because  $L_1 c / c_0 = L$ , where  $c_0$  is the square root of the leading coefficient (which is necessarily positive) of  $N(c)$ .

We shall show that  $a_1$  is in  $F[z]$ . Now  $\bar{a}_1(b_1 + j) - ja_1 = \bar{a}_1 b_1$  is in  $L_1$ ; therefore  $\bar{a}_1 b_1 \equiv 0 \pmod{a_1}$ . Let  $a_1 = a' a''$  where  $a'$  is the largest factor of  $a_1$  in  $F[z]$ ; i.e., the factors of  $a''$  divide no linear polynomials in  $F[z]$ . Then  $b_1 \equiv 0 \pmod{a''}$ . Also in  $L_1$  is  $N(b_1 + j) = b_1 \bar{b}_1 - \sigma$ ; hence  $b_1 \bar{b}_1 - \sigma \equiv 0 \pmod{a_1}$ ,  $\sigma \equiv 0 \pmod{a''}$ . But  $\sigma$  is a product of linear factors in  $F[z]$ ; hence  $a''$  is in  $F$  and  $a_1$  is in  $F[z]$ . We can take the leading coefficient of  $a_1$  to be unity. Then  $L_1$  has the regular basis  $a_1, b_1 + j$  and  $L_1$  is regular. Also  $L$ , being equivalent to  $L_1$ , is likewise regular.

Thus, using Theorem 20, we can always determine whether an integral set of a quaternion algebra  $Q$  having as  $F$  the field of all real numbers has ideals which are not principal.

UNIVERSITY OF CHICAGO,  
CHICAGO, ILL.

## ORDER TYPES AND STRUCTURE OF ORDERS

BY

ANDRÉ GLEYZAL<sup>(1)</sup>

1. **Introduction.** This paper is concerned with operations on order types or order properties  $\alpha$  and the construction of order types related to  $\alpha$ . The reference throughout is to simply or linearly ordered sets, and we shall speak of  $\alpha$  as either property or type. Let  $\alpha$  and  $\beta$  be any two order types. An order  $A$  will be said to be of type  $\alpha\beta$  if it is the sum of  $\beta$ -orders (orders of type  $\beta$ ) over an  $\alpha$ -order; i.e., if  $A$  permits of decomposition into nonoverlapping segments each of order type  $\beta$ , the segments themselves forming an order of type  $\alpha$ . We have thus associated with every pair of order types  $\alpha$  and  $\beta$  the *product order type*  $\alpha\beta$ .

The definition of product for order types automatically associates with every order type  $\alpha$  the order types  $\alpha\alpha = \alpha^2$ ,  $\alpha\alpha^2 = \alpha^3$ ,  $\dots$ . We may furthermore define, for all ordinals  $\lambda$ , a  $\lambda$ th power of  $\alpha$ ,  $\alpha^\lambda$ , and finally a *limit* order type  $\alpha^I$ . This order type has certain interesting properties. It has closure with respect to the product operation, for the sum of  $\alpha^I$ -orders over an  $\alpha^I$ -order is an  $\alpha^I$ -order, i.e.,  $\alpha^I\alpha^I = \alpha^I$ . For this reason we call  $\alpha^I$  *iterative*. In general, we term an order type  $\beta$  having the property that  $\beta\beta = \beta$  *iterative*.  $\alpha^I$  has the following postulational identification:

1.  $\alpha^I$  is a supertype of  $\alpha$ ; that is to say, all  $\alpha$ -orders are  $\alpha^I$ -orders.
2.  $\alpha^I$  is iterative.
3.  $\alpha^I$  is minimal in the sense that all iterative superotypes of  $\alpha$  are superotypes of  $\alpha^I$ .

It may be shown that these conditions determine a unique  $\alpha^I$ , once  $\alpha$  is given. Accordingly, we term  $\alpha^I$  *the minimal iterative supertype of  $\alpha$* . In particular, when we prescribe  $\alpha$  to be the type, "either normal or reverse normal,"  $\alpha^I$ , it turns out, is the type *scattered*<sup>(2)</sup>. Thus we find that scattered orders are constructible from normal and reverse normal orders by the product operation.

Other fundamental operations on orders, such as taking a segment of an order, summing over a normal order, or forming a suborder or superorder of an order, lead to the definition of other order types associated with  $\alpha$ , and to other properties of order types such as *descending*, *extensive*, etc. We denote these associated order types by  $\alpha^D$ ,  $\alpha^E$ ,  $\alpha^F$ ,  $\alpha^R$ , and  $\alpha^N$ . They are unique, de-

Presented to the Society, December 27, 1939, under the title *A general theorem on the structure of linear orders*; received by the editors January 9, 1940.

<sup>(1)</sup> I wish to express my appreciation to Professor H. Blumberg for his generous aid in the preparation of this paper. A summary of its principal results is contained in the Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 291-292.

<sup>(2)</sup> An order is said to be *scattered* if it contains no dense suborders.

pending only upon the choice of  $\alpha$ , and the first four of them have closure and minimal properties analogous to those described for  $\alpha'$ . The type  $\alpha^D$  is shown to be a descending, and  $\alpha^E$  an extensive order type. Also associated with  $\alpha$  are two types which we term  $\alpha$ -dense and  $\alpha$ -scattered. They are, as the names indicate, generalizations of dense and scattered.  $\alpha$ -scattered is iterative and has a certain minimal property with respect to  $\alpha$ . Of particular interest is the case where  $\alpha$  is chosen to be the property of containing  $\aleph_\lambda$  or more elements, where  $\aleph_\lambda$  is the  $\lambda$ th transfinite cardinal. We denote the two order types associated with this  $\alpha$  by  $\aleph_\lambda$ -dense and  $\aleph_\lambda$ -scattered, respectively.  $\aleph_0$ -dense and  $\aleph_0$ -scattered become the properties dense and scattered themselves.

On the basis of these associated order types there may be developed what amounts to an algebra of order types. For example, we may form, by combination, such types as  $\alpha^{DE}$  (meaning  $\beta^E$ , where  $\beta = \alpha^D$ ). The property  $\alpha^{DEI}$ , important in our considerations, is denoted more simply by  $\alpha^T$ . We find, remarkably, that  $\alpha$ -scattered is equivalent to  $\alpha^{ENDEI}$  ( $= \alpha^{ENT}$ ).

A principal result is the following one which gives a decomposition of every order with respect to every order type. It may be stated as follows. *If  $A$  is any order and  $\alpha$  any order type,  $A$  is either of type  $\alpha^T$  or is the sum of  $\alpha^T$ -orders over an order no proper segment<sup>(3)</sup> of which has type  $\alpha^T$ .* This is a generalization of the well known theorem—due to Hausdorff<sup>(4)</sup>—that every order is either scattered or the sum of scattered orders over a dense order. In this paper, the latter decomposition is the one associated with the property “normal or reverse normal.” The order type  $\alpha^T$ , it is found, is simultaneously descending, extensive and iterative. Such a property we term *transitive*. It may be shown, furthermore, that  $\alpha^T$  is the minimal transitive supertype of  $\alpha$ . The property  $\alpha$ -scattered is transitive for all  $\alpha$ , and all transitive order types are supertypes of the type scattered. If  $\alpha$  is transitive,  $\alpha^T$  is equivalent to  $\alpha$  and the above decomposition theorem implies: *If  $A$  is any order and  $\alpha$  a transitive order type,  $A$  is either of type  $\alpha$  or the sum of  $\alpha$ -orders over an order no proper segment of which is of type  $\alpha$ .*

The decompositions we obtain, corresponding to various particular  $\alpha$ 's, give insight into the structure of orders and suggest a number of theorems of general nature. One such theorem we prove is that every order of regular<sup>(5)</sup> cardinal  $\aleph_\lambda$  contains either an  $\aleph_\lambda$ -dense order, or the normal order  $\omega_\lambda$ , or the reverse of  $\omega_\lambda$ .

An order  $I_\lambda$  of transfinite integers is introduced which satisfies the following universality conditions:  $I_\lambda$  is scattered, of cardinal  $\aleph_\lambda$ , and contains all scattered orders of cardinal less than  $\aleph_\lambda$ .

Problems arise as to properties and methods for constructing orders of

<sup>(3)</sup> By a proper segment we understand a segment with more than one element.

<sup>(4)</sup> *Grundsätze der Mengenlehre*, pp. 95–97.

<sup>(5)</sup> The cardinal  $\aleph_\lambda$  and the normal order  $\omega_\lambda$  initiating the cardinal  $\aleph_\lambda$  are said to be regular if every suborder of  $\omega_\lambda$  cofinal with  $\omega_\lambda$  is of type  $\omega_\lambda$ .

types such as  $\aleph_\lambda$ -scattered or  $\omega_\lambda$ -scattered. The considerations of this paper lead also to other problems on orders and order types. A number of these are alluded to (see §12), their solution awaiting future research.

**2. Decomposition of an order.** Let  $A$  be a given order and  $\alpha$  an order type or order property. The problem which we wish to consider is that of obtaining a composition of  $A$  in terms of orders of type  $\alpha$ . Later, we give a formal proof of the composition theorem stated in the introduction, but we shall first proceed inductively, tracing step-by-step, the ideas leading to the result we have in mind. To obtain a segmental decomposition, i.e., a separation of the order  $A$  into segments, let us begin by associating with an element  $a$  of  $A$  the elements  $e$  of  $A$  such that the segment  $(a, e)$  or  $(e, a)$ —taken to include end-elements—has property  $\alpha$ . The elements thus associated with  $a$  form a set  $S_a$  which may or may not be a segment of  $A$ . To insure that  $S_a$  form a segment, let us require that  $\alpha$  be such that every initial and every final segment of an  $\alpha$ -order (an order having property  $\alpha$ ) be likewise an  $\alpha$ -order. This is equivalent to requiring that every segment of an  $\alpha$ -order be an  $\alpha$ -order. This condition upon  $\alpha$  we express by saying that  $\alpha$  is a *descending* order property. Furthermore, to insure that two different sets  $S_a$  have no common elements, we ask that the sum<sup>(\*)</sup> of two  $\alpha$ -orders be again an  $\alpha$ -order. An order property obeying the latter condition we term *additive*. Therefore, if  $\alpha$  is a descending, additive order property, the order  $A$  is the sum of nonoverlapping segments  $S_a$  as defined, and we have determined a composition for  $A$ . The segments  $S_a$  themselves form a new order  $A_1$  if we set  $S_a < S_{a'}$  when  $a < a'$ , and we shall say that  $A$  is the sum of  $\alpha$ -orders over the base  $A_1$ . In the same way as  $A$ , the order  $A_1$  may be decomposed with respect to  $\alpha$ , yielding a new base order  $A_2$  whose elements are now segments of  $A_1$ . Since each element of  $A_1$  is a segment of  $A$ , we may again consider the elements of  $A_2$  as segments of  $A$ . Continuing, we secure, for every integer  $n$ , the base order  $A_n$  with elements interpretable as segments of  $A$ . A "limit" order  $A_\omega$  may be formed as follows. Let  $a$  be an arbitrary element of  $A$ , and  $S_a$  the set of elements  $e$  of  $A$  belonging, for some  $n$ , to the elements of  $A_n$  containing  $a$ . The set  $S_a$  is a segment of  $A$  and the sum of such segments constitutes  $A$ . Let  $A_\omega$  be the order with these segments as elements. We may now continue this process beyond the  $\omega$ th stage until finally we reach an order  $A_\mu$ , where  $\mu$  is a transfinite ordinal, such that no proper segment of  $A_\mu$  has property  $\alpha$ . The order  $A$  is the sum of segments  $S$  of  $A$  over the base order  $A_\mu$ . We observe that each segment  $S$  may be built up from  $\alpha$ -orders by means of the following operations:

- (1) Forming an order by substituting  $\alpha$ -orders for the elements of an  $\alpha$ -order or an order already constructed.

(\*) By the *sum*  $A+B$  of two orders  $A$  and  $B$  is meant the order formed by placing all elements of  $B$  after all elements of  $A$ , no change being made in the relative position of elements in  $A$  or in  $B$ .

(2) Forming an order by substituting  $\alpha$ -orders or orders already constructed for the elements of a normal or reverse normal order.

Let us call an order which may be built up from  $\alpha$ -orders by means of these two operations an  $\alpha^T$ -order. We may then say that  $A$  is the sum of  $\alpha^T$ -orders over a base order no proper segment of which has property  $\alpha$ . It may be furthermore shown that  $A$  has no proper segment with property  $\alpha^T$ , but we defer the proof until later. We choose to start anew making use of the notions we have just obtained.

**3. Iterative order type.** Let  $\alpha$  and  $\beta$  be any two given order properties or types, for example, *perfect* and *scattered*. We say that an order is of property  $\alpha\beta$  if it is the sum of  $\beta$ -orders over an  $\alpha$ -order. As stated in §1, we term an order property  $\alpha$  *iterative* if it has "closure" with respect to the operation of summing over itself; i.e., if the sum of  $\alpha$ -orders over an  $\alpha$ -order is again an  $\alpha$ -order.

Suppose  $\alpha$  is not an iterative order property. We may construct an iterative property  $\beta$  implied by  $\alpha$ , as follows: By  $\alpha^1$  we understand  $\alpha$  itself. Suppose  $\alpha^\mu$  is defined for all ordinals  $\mu$  less than  $\lambda$ . An order  $A$  will be said to have property  $\alpha^\lambda$  if it is the sum of  $\alpha^\mu$ -orders over an  $\alpha$ -order, where  $\mu < \lambda$  and  $\mu$  is permissibly variable.  $\beta$  is then the *sum type* of all types(?)  $\alpha^\lambda$ ; i.e., an order  $A$  will be said to have property  $\beta$  if it has property  $\alpha^\lambda$  for some  $\lambda$ . We shall denote the property  $\beta$  associated in this way with  $\alpha$  by  $\alpha^I$ , the superscript  $I$  signifying that  $\alpha^I$  is iterative, as we show later<sup>(8)</sup>.

We prove, for future reference, that  $\alpha^{\mu+\lambda}$  is a supertype of  $\alpha^\lambda\alpha^\mu$ . Let us denote by  $\alpha^\mu$  that type which is the sum type of all types  $\alpha^\nu$ , where  $\nu < \mu$ . Our definition of  $\alpha^\mu$  may then be written  $\alpha^\mu = \alpha\alpha^\mu$ . We may then write  $\alpha^{\mu+1} = \alpha\alpha^{\mu+1} = \alpha(\alpha^\mu + \alpha^\mu) = \alpha(\alpha^\mu + \alpha^\mu)$ . Hence  $\alpha^{\mu+1}$  is a supertype of  $\alpha\alpha^\mu$  and the statement holds for  $\lambda = 1$ . Suppose it holds for all ordinals less than  $\lambda$ . We may write  $\alpha^{\mu+\lambda} = \alpha\alpha^{\mu+\lambda}$ . Our hypothesis implies, however, that  $\alpha^{\mu+\lambda}$  is a supertype of  $\alpha^\lambda\alpha^\mu$ . Therefore  $\alpha^{\mu+\lambda} = \alpha\alpha^{\mu+\lambda}$  is a supertype of  $\alpha\alpha^\lambda\alpha^\mu = \alpha^\lambda\alpha^\mu$ <sup>(9)</sup>. In particular, we note that  $\alpha^{1+\omega}$  is a supertype of  $\alpha^\omega\alpha = (\alpha\alpha^\omega)\alpha = \alpha(\alpha^\omega\alpha) = \alpha\alpha^\omega = \alpha^\omega$ , as would also be expected from the relation  $1+\omega = \omega$ . We show that  $\alpha^I$  has certain minimal and uniqueness properties in relation to  $\alpha$ , and that these provide a postulational definition for  $\alpha^I$ .

We introduce, for sets in general, a notion of a *minimal property*. Let  $\alpha$  stand for a given set property, and  $A$  for a given property of set properties.

(?) Let  $S$  be a set of order types  $\alpha$ . By the *sum type* of the order types of  $S$  we understand the order type  $\beta$  defined as follows. An order will be said to be of type  $\beta$  if it has property  $\alpha$  for some  $\alpha$  of  $S$ ; otherwise, it will be said not to have type  $\beta$ .

(8) It may be true that for a given  $\alpha$  there always exists a first ordinal  $\lambda$  such that  $\alpha^\lambda$  is iterative, but the author has no proof of this.

(9) The *exponential law*  $\alpha^{\mu+\lambda} = \alpha^\lambda\alpha^\mu$  holds if the order consisting of a single element has type  $\alpha$ . For  $\alpha^\mu$  is then a supertype of  $\alpha^\nu$  for all  $\nu < \mu$  and, consequently,  $\alpha^{\mu+1} = \alpha\alpha^{\mu+1} = \alpha\alpha^\mu$ . Assuming  $\alpha^{\mu+\nu} = \alpha^\nu\alpha^\mu$  for all  $\nu < \lambda$ , we have  $\alpha^{\mu+\lambda} = \alpha\alpha^{\mu+\lambda} = \alpha\alpha^\lambda\alpha^\mu = \alpha^\lambda\alpha^\mu$ .



The set property  $\beta$  will be said to be a *minimal  $A$ -property implied by  $\alpha$* , if it is implied by  $\alpha$ , has property  $A$ , and is such that if  $\beta'$  is any set property implied by  $\alpha$  and having property  $A$ , it is implied by  $\beta$ . Two minimal  $A$ -properties implying  $\alpha$  are equivalent in the sense that each implies the other. We may thus regard the minimal  $A$ -property implied by  $\alpha$  as uniquely determined—if it exists. We shall therefore speak of “the” instead of “a” minimal property.

**THEOREM 1.**  $\alpha^I$  is the minimal iterative property implied by  $\alpha$ .

**Proof.** Suppose an order  $A$  is a sum of  $\alpha^I$ -orders  $A_\lambda$  over an  $\alpha^I$ -order  $\Lambda = \{\lambda\}$ , the subscript  $\lambda$  ranging over all the elements of the order  $\Lambda$ . Each  $A_\lambda$  is an  $\alpha^\mu$ -order for some ordinal  $\mu$ . We set  $\sigma$  equal to the first ordinal larger than any of the ordinals  $\mu$ . If  $\Lambda$  is an  $\alpha^I$ -order,  $A$  is an  $\alpha^\sigma$ -order. Therefore  $A$  is an  $\alpha^{\sigma+\sigma}$ -order and hence an  $\alpha^I$ -order.  $\alpha^I$  is therefore iterative. Now let  $\beta$  be an iterative property implied by  $\alpha$ . Assume  $\beta$  is implied by  $\alpha^\mu$  for  $\mu < \lambda$ .  $\beta$  is then implied by  $\alpha^\lambda$  since  $\beta$  is iterative. Consequently  $\beta$  is implied by  $\alpha^\lambda$  for all ordinals  $\lambda$ , that is, by  $\alpha^I$ ; and the theorem is true.

There is thus uniquely associated with every  $\alpha$  the property  $\alpha^I$  which is the minimal iterative property implied by  $\alpha$ ; i.e., the minimal iterative type which includes  $\alpha$  as subtype. We shall say alternatively, that  $\alpha^I$  is the minimal iterative supertype of  $\alpha$ . The latter phrasing will also be employed, when convenient, for order type properties other than iterative.

**4. Descending order type.** We consider now the property *descending* for order types. Let us denote by  $\alpha^D$  the property of being a segment of an  $\alpha$ -order. A segment of an  $\alpha^D$ -order is a segment of an  $\alpha$ -order and consequently an  $\alpha^D$ -order. Thus  $\alpha^D$  is descending. If  $\beta$  is a descending property implied by  $\alpha$ , every segment of every  $\alpha$ -order is a  $\beta$ -order and  $\beta$  is implied by  $\alpha^D$ . Accordingly, we may state

**THEOREM 2.**  $\alpha^D$  is the minimal descending supertype of  $\alpha$ .

One may ask the nature of the properties  $\alpha^{DI}$  ( $= \beta^I$ , where  $\beta = \alpha^D$ ), or  $\alpha^{ID}$ , etc., composed by combining the above described processes. We find that

**THEOREM 3.**  $\alpha^{DI}$  is the minimal descending and iterative supertype of  $\alpha$ .

**Proof.**  $\alpha^D = (\alpha^D)^1$  is descending, as we have seen. Suppose  $(\alpha^D)^\mu$ , for ordinals  $\mu < \lambda$ , is descending. If an order  $A$  has property  $(\alpha^D)^\lambda$ , it is the sum of  $(\alpha^D)^\mu$ -orders,  $\mu < \lambda$ , over an  $\alpha^D$ -order, and every segment  $S$  of  $A$  is the sum of segments of  $(\alpha^D)^\mu$ -orders over a segment of an  $\alpha^D$ -order. Hence  $S$  is the sum of  $(\alpha^D)^\mu$ -orders over an  $\alpha^D$ -order and has property  $(\alpha^D)^\lambda$ . Therefore  $(\alpha^D)^\lambda$  is descending for all ordinals  $\lambda$  and it follows that  $\alpha^{DI}$  is descending. By Theorem 1,  $\alpha^{DI}$  is iterative. Suppose now  $\beta$  is a descending and iterative property implied by  $\alpha$ . Then  $\beta$  is implied by  $\alpha^D$ , and therefore by  $\alpha^{DI}$ , for  $\alpha^D$  and  $\alpha^{DI}$  are minimal. Thus  $\alpha^{DI}$  is descending and iterative and is minimal, as was to be proved.

5. **Extensive order type.** We introduce a third property—again a closure property—corresponding to the operation described above (§2), of summing orders over normal orders or reverse normal orders. An order type  $\alpha$  will be termed *extensive* if the sum of  $\alpha$ -orders over a normal order or a reverse normal order is an  $\alpha$ -order. We prove later that the minimal extensive supertype of  $\alpha$  exists for all  $\alpha$ , and is equivalent to  $\sigma\alpha$ , where  $\sigma$  is the property *scattered*. In conformity with the notation previously employed for supertypes of  $\alpha$ , we shall denote  $\sigma\alpha$  by  $\alpha^E$ .

6. **Transitive order type.** The constructions of  $\alpha^D$ ,  $\alpha^I$  and  $\alpha^E$  are based on three operations described as follows.  $\alpha^D$  has closure with respect to the operation of taking segments, for a segment of an  $\alpha^D$ -order is an  $\alpha^D$ -order. Accordingly, we term this operation a *D*-operation. Similarly,  $\alpha^I$  has closure with respect to the *I*-operation of summing orders of a certain type over orders of the same type. Also,  $\alpha^E$  has closure with respect to the *E*-operation of summing over normal or reverse normal orders. We now define an order type having closure with respect to all three of the above operations. An order property will be said to be *transitive* if it is iterative, descending and extensive. Later it is shown that the minimal transitive supertype of  $\alpha$  exists and is the order type  $(\sigma\alpha^D)^I$ .

If  $\alpha$  is descending, a single element has property  $\alpha$ . Consequently a descending and extensive order type includes normal and reverse normal orders as subtype. Later, we shall see that a transitive order type includes scattered as subtype.

7. **Decomposition of an order into  $\alpha$ -orders.** We prove now the following fundamental decomposition theorem:

**THEOREM 4a.** *If  $\alpha$  is a descending and iterative order property, and  $A$  an order whose normal orders and reverse normal orders have property  $\alpha$ , then  $A$  has property  $\alpha$  or is the sum of  $\alpha$ -orders over an order no proper segment of which has property  $\alpha$ .*

**Proof.** In the special case where  $A$  consists of exactly one element, the theorem is true. Suppose  $A$  has more than one element. We shall say that a segment of  $A$  is a maximal  $\alpha$ -segment if it has property  $\alpha$  and no segment properly containing it has property  $\alpha$ . Every element  $a$  of  $A$  is contained in a maximal  $\alpha$ -segment. For let  $S_a$  be the set of elements  $e$  such that the segment  $\langle a, e \rangle$  or  $\langle e, a \rangle$  has property  $\alpha$ , the symbol  $\langle \rangle$  signifying that the end points of the segment are included. The set  $S_a$  is a segment of  $A$ . For if  $e'$  is an element of  $A$  between  $a$  and an element  $e$  of  $S_a$ ,  $\langle a, e' \rangle$  or  $\langle e', a \rangle$  is a segment of  $\langle a, e \rangle$  or  $\langle e, a \rangle$  respectively, and, since  $\alpha$  is descending, has property  $\alpha$ . Thus  $e'$  is an element of  $S_a$ . The segment  $S_a$  has property  $\alpha$ . For let  $a = e_1, e_2, \dots; \dots, e_\lambda, \dots$  be a normal suborder of  $S_a$  cofinal with  $S_a$ . A segment  $\langle e_\lambda, e_{\lambda+1} \rangle$ —taken to include  $e_{\lambda+1}$  but not  $e_\lambda$ —is an  $\alpha$ -order, since  $\langle a, e_{\lambda+1} \rangle$  is an  $\alpha$ -order. By hypothesis, the normal order  $e_1, e_2, \dots; \dots, e_\lambda, \dots$

is an  $\alpha$ -order. Thus the suborder of elements of  $S_a$  to the right of  $a$  is the sum of  $\alpha$ -orders over an  $\alpha$ -order and therefore has property  $\alpha$ . Similarly, the suborder of  $S_a$  to the left of  $a$  is an  $\alpha$ -order. Thus, the segment  $S_a$  is the sum of at most three  $\alpha$ -orders and is consequently an  $\alpha$ -order, since, by hypothesis, every finite suborder of  $A$  is an  $\alpha$ -order. From our definition of  $S_a$ , it follows that no segment properly containing  $S_a$  has property  $\alpha$ , and  $S_a$  is hence a maximal  $\alpha$ -segment. Moreover, no two distinct maximal  $\alpha$ -segments have elements in common, for the sum of two such segments forms an  $\alpha$ -order properly containing each of them, contrary to the definitional property of the maximal  $\alpha$ -segment. Let  $B$  be the order consisting of these maximal  $\alpha$ -segments. We have shown that  $A$  is the sum of maximal  $\alpha$ -segments over the order  $B$ . No proper segment of  $B$  has property  $\alpha$ . For suppose there exists such a segment. Then there exists a subsegment  $\langle S_a, S_b \rangle$ ,  $S_a \neq S_b$ , of  $B$  which has property  $\alpha$ . Since  $\alpha$  is iterative, the set of elements of  $A$  composing the segment  $\langle S_a, S_b \rangle$  has property  $\alpha$ . It follows that  $b$  is an element of  $S_a$  and consequently that  $S_a = S_b$ , contrary to hypothesis. The theorem is thus proved.

**THEOREM 4b.** *If  $\alpha$  is a transitive order property and  $A$  a given order, then  $A$  either has property  $\alpha$  or is the sum of  $\alpha$ -orders over an order no proper segment of which has property  $\alpha$ .*

**Proof.** We have seen that  $\alpha$  includes the order types normal and reverse normal as subtypes. In particular, all normal and reverse normal orders contained by  $A$  as suborders are  $\alpha$ -orders, and Theorem 4a applies.

**THEOREM 4c.** *If  $\alpha$  is a transitive order property and  $A$  a given order, then  $A$  either has property  $\alpha$  or is the sum of  $\alpha$ -orders over a dense order.*

**Proof.** Suppose  $A$  is not an  $\alpha$ -order. Then, by Theorem 4b,  $A$  is the sum of  $\alpha$ -orders over an order  $B$  no proper segment of which has property  $\alpha$ . Every proper segment of  $B$  consequently does not consist of a finite number of elements, since  $\alpha$  includes finite order types. We conclude  $B$  is a dense order and the theorem follows.

**8. Properties of the type scattered.** By means of the above decompositions, we prove a number of theorems concerning the type *scattered*. For convenience, we shall denote this type by the symbol  $\sigma$ .

**THEOREM 5.** *The order type scattered is transitive.*

**Proof.** For suppose an order  $A$  has property  $\sigma$ . Then it contains no dense suborders and surely no segment of  $A$  contains a dense suborder. Consequently,  $\sigma$  is a descending order property.  $\sigma$  is iterative; for let  $A$  be an order which is the sum of  $\sigma$ -orders  $A_\lambda$  over a  $\sigma$ -order  $\Lambda = \{\lambda\}$ . Assume there exists a dense suborder  $D$  of  $A$ . No  $A_\lambda$  contains more than one element of  $D$  since otherwise  $A_\lambda$  would contain a dense order. Thus  $D$  is a dense suborder of  $\Lambda$ , contrary to hypothesis.  $A$  therefore is scattered and consequently  $\sigma$  is an

iterative order property. Furthermore, normal and reverse normal orders are subtypes of scattered orders and we conclude that  $\sigma$  is transitive. Substituting  $\sigma$  for  $\alpha$  in the decomposition theorem, there results

**THEOREM 6.** *Every order is either scattered or the sum of scattered orders over a dense order<sup>(10)</sup>.*

**THEOREM 7.** *Every transitive order type includes the type scattered as subtype.*

**Proof.** Let  $\alpha$  be a transitive order type and  $A$  any scattered order. By Theorem 4c,  $A$  has either property  $\alpha$  or is the sum of  $\alpha$ -orders over a dense order. In the latter case  $A$  would contain a dense suborder, contrary to hypothesis. Therefore  $A$  has property  $\alpha$  and the theorem is proved.

**9. Minimal supertypes.** We now prove the following theorem:

**THEOREM 8.** *The minimal iterative order type which includes normal and reverse normal as subtypes is the type scattered.*

**Proof.** Let  $\alpha$  be the property of being either a normal or reverse normal order. We have seen (Theorem 3) that  $\alpha^{DI}$  is descending and iterative. But  $\alpha^D = \alpha$ , and consequently  $\alpha^{DI} = \alpha^I$ . Thus the minimal iterative property  $\alpha^I$  implied by  $\alpha$  exists and is descending.  $\alpha^I$  is surely extensive since it is iterative and includes normal and reverse normal as subtypes. Therefore  $\alpha^I$  is transitive and hence contains  $\sigma$  as subtype. But, since  $\sigma$  is transitive, it is iterative, and must therefore include the minimal type  $\alpha^I$ . Thus  $\alpha^I$  is equivalent to  $\sigma$  and the theorem is valid.

**THEOREM 9.** *If  $\alpha$  is any order type, the minimal extensive supertype of  $\alpha$  is  $\sigma\alpha$ , where  $\sigma$  is the order type scattered.*

**Proof.** Let  $\beta$  be the property normal or reverse normal. Every  $\beta\sigma$ -order is a  $\sigma^2$ -order, hence a  $\sigma$ -order. Thus  $\beta\sigma\alpha = \sigma\alpha$ , and  $\sigma\alpha$  is extensive. Let now  $\gamma$  be any extensive order type including  $\alpha$  as subtype.  $\gamma$  then includes  $\beta\alpha$ , hence includes  $\beta\beta\alpha = \beta^2\alpha$ ,  $\beta^3\alpha$ , etc. Suppose  $\gamma$  includes  $\beta^\mu\alpha$  as subtype for  $\mu < \lambda$ . It then includes the sum of  $\beta^\mu\alpha$ -orders,  $\mu < \lambda$ , over a  $\beta$ -order; that is,  $\beta^\lambda\alpha$  as subtype. Thus  $\gamma$  includes  $\beta^\lambda\alpha$  for all ordinals  $\lambda$ . Therefore  $\gamma$  includes  $\beta^I\alpha$  as subtype, and since, by Theorem 8,  $\beta^I = \sigma$ , we conclude that  $\sigma\alpha$  is minimal as stated in the theorem.

**THEOREM 10.** *If  $\alpha$  is any order type, the minimal transitive supertype of  $\alpha$  is the type  $(\sigma\alpha^D)^I = \alpha^{D\beta I}$ .*

**Proof.** In general, the product of two descending order types is a descending order type, and we infer that  $\sigma\alpha^D$  is descending. By Theorem 3, then,  $(\sigma\alpha^D)^I$  is descending and iterative. Clearly,  $(\sigma\alpha^D)^I$  includes  $\sigma$ , and hence in-

<sup>(10)</sup> Cf. Hausdorff, loc. cit.

cludes normal and reverse normal as subtype. Hence, since  $(\sigma\alpha^D)^I$  is iterative, it is extensive. Consequently,  $(\sigma\alpha^D)^I$  is transitive. Now, let  $\beta$  be any transitive type which includes  $\alpha$  as subtype.  $\beta$  is descending and includes therefore the minimal property  $\alpha^D$ . Likewise,  $\beta$  is extensive and includes therefore  $\sigma\alpha^D$ . Again, since  $\beta$  is iterative, it includes  $(\sigma\alpha^D)^I$ . Thus  $(\sigma\alpha^D)^I$  is the minimal transitive order type called for in the theorem.

We denote, for brevity, the transitive property  $(\sigma\alpha^D)^I$  associated with  $\alpha$ , by  $\alpha^T$ . Combining Theorems 4b and 10, we may state

**THEOREM 11.** *If  $A$  is any order, and  $\alpha$  any given order property, either  $A$  has property  $\alpha^T$  or is the sum of  $\alpha^T$ -orders over an order no proper segment of which has property  $\alpha^T$ , where  $\alpha^T$  is the minimal transitive property implied by  $\alpha$ .*

We now determine minimal properties for a number of particular order properties. Instead of speaking of an order property we will find it convenient, on occasion, to speak of the set of orders having the property. We introduce, for this purpose, the following terminology. Let  $S$  be a given set of orders, and  $\sigma$  the property of belonging to  $S$ . We understand by the *minimal, iterative set containing  $S$*  the set  $M$  of orders such that the property of belonging to  $M$  is equivalent to the minimal iterative property implied by  $\sigma$ . An analogous phrasing will be used for order type properties other than iterative.

**THEOREM 12.** *The minimal iterative and descending set containing the set which consists of the single element  $\omega_\lambda$  ( $\omega_\lambda$ -reversed) is the set of all normal orders (reverse normal orders) of cardinal  $\aleph_\lambda$  or less.*

**Proof.** Clearly, the set of all normal orders of cardinal  $\aleph_\lambda$  or less is iterative and descending. Suppose all proper initial segments of  $\nu$ , where  $\nu$  is a normal order of cardinal  $\aleph_\lambda$  belong to  $M$ , the minimal iterative and descending set containing  $\omega_\lambda$ . If  $\nu$  has a last element,  $\nu$  itself is an element of  $M$  since the normal orders  $\nu - 1$ , 1 and 2 are in  $M$ . If  $\nu$  has no last element, there exists a normal suborder of ordinals  $\nu_1, \nu_2, \dots; \nu_\sigma, \dots$  of  $\nu$ , cofinal with  $\nu$  and of type  $\omega_\lambda$  or less. Thus  $\nu$  is an initial segment of the normal order whose ordinal is  $\nu_1 + \nu_2 + \dots; + \dots + \nu_\sigma + \dots$ , with  $\nu_\sigma$  in  $M$ . Consequently  $\nu$  is an element of  $M$ —similarly for reverse normal orders.

**THEOREM 13.** *The minimal iterative and descending set containing the set consisting of the two elements  $\omega_\lambda, \omega_\lambda$ -reversed, is the set of all scattered orders of cardinal  $\aleph_\lambda$  or less.*

**Proof.** Clearly, the set of scattered orders of cardinal  $\aleph_\lambda$  or less is iterative and descending. Now, let  $\alpha$  be any iterative and descending order type which includes the type  $\omega_\lambda$  and  $\omega_\lambda$ -reversed, and let  $A$  be any scattered order of cardinal  $\aleph_\lambda$  or less. By Theorem 12, all normal orders and reverse normal orders of  $A$  have property  $\alpha$ . But by Theorem 4a,  $A$  is either an  $\alpha$ -order or is the sum of  $\alpha$ -orders over a dense order. In the latter case  $A$  would contain a



dense order, contrary to hypothesis. Thus the scattered orders of cardinal  $\aleph_\lambda$  have property  $\alpha$  and constitute the minimal set described in the theorem.

We choose, thirdly, for a particular set of orders, the set  $S$  of orders of cardinal less than  $\aleph_\lambda$ . We have seen that by means of the operations of taking segments and summing over certain orders, we may construct the orders of the set  $M$  which is the minimal transitive set containing  $S$  as subset. We inquire now as to the nature of an order of  $M$ . The decomposition theorem shows us that an order  $A$  is either an order of  $M$  or is the sum of orders belonging to  $M$  over an order  $B$  no proper segment of which belongs to  $M$ . Suppose the latter is true. Then no proper segment of  $B$  is of cardinal less than  $\aleph_\lambda$  and consequently every proper segment of  $B$  is of cardinal  $\aleph_\lambda$  or more. This property of  $B$  suggests the notion  $\aleph_\lambda$ -dense which we define as follows: An order will be said to be  $\aleph_\lambda$ -dense if it has more than one element and every proper segment of it contains  $\aleph_\lambda$  or more elements. Thus  $\aleph_\lambda$ -dense is a generalization of the property dense,  $\aleph_0$ -dense being equivalent to the property dense.

Let us return to the consideration of the properties of an order of  $M$ . We have found that  $A$  is either in  $M$  or contains an  $\aleph_\lambda$ -dense suborder. For the purpose of insuring that  $A$  be in  $M$  we need merely specify that no suborder of  $A$  be  $\aleph_\lambda$ -dense.  $A$  will then be said to be  $\aleph_\lambda$ -scattered, and in general an order possessing no  $\aleph_\lambda$ -dense suborder will be termed  $\aleph_\lambda$ -scattered. Thus  $\aleph_\lambda$ -scattered is a generalization of the property scattered, the latter being equivalent to  $\aleph_0$ -scattered. Making the guess that, conversely, all orders of  $M$  are  $\aleph_\lambda$ -scattered we venture

**THEOREM 14.** *The minimal transitive set which contains the set of all orders of cardinal  $\aleph_\lambda$  is the set of  $\aleph_\lambda$ -scattered orders.*

**Proof.** We have seen that the set  $A$  of  $\aleph_\lambda$ -scattered orders is a subset of  $M$ , the minimal transitive set containing all orders of cardinal less than  $\aleph_\lambda$ . We show conversely, that  $M$  is a subset of  $A$ . By Theorem 5, the set of  $\aleph_0$ -scattered orders is a transitive set. If we substitute  $\aleph_\lambda$ -scattered for scattered and  $\aleph_\lambda$ -dense for dense in the proof of this theorem we secure a proof that the property  $\aleph_\lambda$ -scattered is transitive for all  $\lambda$ . Also,  $A$  contains all orders of cardinal less than  $\aleph_\lambda$ . But  $M$ , being a minimal transitive set, is then a subset of the transitive set  $A$ . Thus  $M$  and  $A$  are identical and  $\aleph_\lambda$ -scattered is the required minimal property.

**10. Transfinite integer.** We now define an order  $I$  which is a generalization of the order of the positive and negative integers, and which is a universal scattered order in the sense that it contains all scattered orders as suborders. By *transfinite integer* we understand the form:

$$\sum_{i=1}^n a_i \omega^{\alpha_i} = a_1 \omega^{\alpha_1} + a_2 \omega^{\alpha_2} + \cdots + a_n \omega^{\alpha_n},$$

where the coefficients  $a$  are ordinary integers, and the exponents  $\alpha_i$  are ordinals decreasing as  $i$  increases. We denote the totality of transfinite integers by  $I^{(11)}$ . An element  $\sum_{i=1}^n a_i \omega^{\alpha_i}$  of  $I$  will be said to be less than an element  $\sum_{i=1}^m b_i \omega^{\beta_i}$  of  $I$ , if for the first index at which the two forms disagree either  $\alpha_i < \beta_i$  or  $\alpha_i = \beta_i$  and  $a_i < b_i$ . It is seen that  $I$  thus becomes a linear order. We define *sum* and *product* for two transfinite integers in the customary algebraic manner.

**THEOREM 15a.**  *$I$  is scattered and every scattered order is similar to a suborder of  $I$ .*

**Proof.** Let  $\alpha$  be the property of being a suborder of  $I$  which is not coinitial nor cofinal with  $I$ . Clearly,  $\alpha$  is descending. It is also iterative. For suppose  $A$  is the sum of  $\alpha$ -orders  $A_\lambda$  over an  $\alpha$ -order  $\Lambda = \{\lambda\}$ .  $A$  is then isomorphic with the linear order which consists of the pairs  $(\lambda, a_\lambda)$ , where  $\lambda$  is any element of  $\Lambda$  and  $a$  any element of  $A_\lambda$ , ordered first according to  $\lambda$  and then according to  $a_\lambda$ . Let  $\omega^\alpha$  be the first power of  $\omega$  such that it is greater than, and  $-\omega^\alpha$  less than, every transfinite integer occurring in the orders  $A_\lambda$ . The transfinite integers of the form

$$\sum_{j=1}^m 2b_j \omega^{\alpha+\beta_j} + \sum_{i=1}^n a_i \omega^{\alpha\lambda_i},$$

where  $\sum_{j=1}^m b_j \omega^{\beta_j}$  is an element  $\lambda$  of  $\Lambda$ , and  $\sum_{i=1}^n a_i \omega^{\alpha\lambda_i}$  is an element  $a_\lambda$  of  $A_\lambda$ , constitute a suborder of  $I$  similar to  $A$ , for the correspondence

$$\sum_{j=1}^m 2b_j \omega^{\alpha+\beta_j} + \sum_{i=1}^n a_i \omega^{\alpha\lambda_i} \sim (\lambda, a_\lambda)$$

is biunique and preserves order.  $\alpha$  is thus descending and iterative. Suppose now  $A$  is a scattered order. It is either an  $\alpha$ -order or the sum of  $\alpha$ -orders over a dense order, by Theorem 4c. The latter case cannot occur, for  $A$  would then contain a dense order, contrary to hypothesis.  $A$  is therefore a suborder of  $I$ .

We now show, conversely, that every suborder of  $I$  is scattered. Upon "writing out" any segment of the order of the transfinite integers, as for example:  $1, 2, 3, \dots; \dots, \omega-3, \omega-2, \omega-1, \omega, \omega+1, \omega+2, \omega+3, \dots; \dots, 2\omega-3, 2\omega-2, 2\omega-1, 2\omega, 2\omega+1, 2\omega+2, 2\omega+3, \dots; \dots, 3\omega \pm n, \dots; \dots, \omega^2 \pm n\omega \pm m, \dots$ , we see that it is *locally symmetric* in the sense that for every Dedekind cut  $(A, B)$  of  $I$ ,  $A$  and  $B$  both non-null, there exist subsegments  $A_1$  of  $A$  and  $B_1$  of  $B$ , cofinal and coinitial respectively with  $A$  and  $B$ , such that  $A_1$  is similar to  $B_1$  reversed<sup>(12)</sup>. Suppose  $I$  is not a scattered order.

<sup>(11)</sup> We introduce  $I$ , rather than a segment of  $I$ , for convenience, not insisting upon the logical character of  $I$  as a totality.

<sup>(12)</sup> Moreover, it is clear that  $I$ , or segments of  $I$ , are the only orders, except for isomorphism, with this local symmetry.

By Theorem 6, it is the sum of scattered orders  $A_\lambda$  over a dense order  $\Lambda = \{\lambda\}$ . No final segment of  $A_\mu$ , where  $\mu$  is a fixed element of  $\Lambda$ , can be similar to the reverse of any initial segment of the set of elements to the right of  $A_\mu$ . For every such segment contains a dense order and  $I$  would then not have local symmetry, contrary to fact. We conclude  $I$  is scattered.

We shall say that a transfinite integer  $\alpha$  is of cardinal  $\aleph_\mu$  if there are  $\aleph_\mu$  elements in the segment  $(0, \alpha)$ . The segment of  $I$  comprised of all elements of cardinal less than  $\aleph_\mu$  we shall indicate by  $I_\mu$ .

**THEOREM 15b.**  $I_\mu$  is scattered and contains as suborder all scattered orders of cardinal less than  $\aleph_\mu$ .

**Proof.** Of course  $I_\mu$ , being a suborder of  $I$ , is scattered. Let us assume, first,  $\aleph_\mu$  is a regular cardinal. Substituting  $I_\mu$  for  $I$  in the proof of Theorem 15a there results a proof that the property  $\alpha$  of being a suborder of  $I_\mu$  not coinitial nor cofinal with  $I_\mu$  is descending and iterative. If  $A$  is a scattered order of cardinal less than  $\aleph_\mu$ , all normal and reverse normal suborders of  $A$  are of cardinal less than  $\aleph_\mu$  and hence have property  $\alpha$ . As above, it follows that  $A$  is a suborder of  $I_\mu$ . The theorem is thus proved for regular cardinals  $\aleph_\mu$ . Suppose, now,  $\aleph_\mu$  is not regular. Then  $\aleph_\mu$  has no cardinal which is an immediate predecessor and is therefore expressible as a sum of regular cardinals  $\aleph_\nu$ , with  $\nu < \mu$ . Consequently, if  $A$  is a scattered order of cardinal less than  $\aleph_\mu$ , we reason  $A$  has cardinal less than some regular cardinal  $\aleph_\nu$ , with  $\nu < \mu$ . Therefore  $A$  is a suborder of  $I_\nu$ . Inasmuch as  $I_\nu$  is a segment of  $I_\mu$ ,  $A$  is a suborder of  $I_\mu$ . The theorem is thus proved for all cardinals  $\aleph_\mu$ .

There are  $2^{\aleph_\lambda}$  scattered orders of cardinal  $\aleph_\lambda$ . Accordingly, one may form an order  $S$  which is scattered and contains as suborders all scattered orders of cardinal less than  $\aleph_1$ , for example, simply by summing all such orders over a normal order of cardinal  $2^{\aleph_0}$ . Let us compare the two orders  $S$  and  $I_1$ , both of which are scattered and contain all scattered orders of cardinal less than  $\aleph_1$ .  $I_1$  is certainly of cardinal  $\aleph_1$  whereas  $S$  is of cardinal  $2^{\aleph_0}$ . Hence, they are both of cardinal  $\aleph_1$ , only if the continuum hypothesis  $\aleph_1 = 2^{\aleph_0}$  is true.

Let us now compare  $I_\mu$  to the order  $S$  composed by summing all scattered orders of cardinal less than  $\aleph_\mu$  over a normal order. This normal order must be of cardinal  $\aleph_\nu = \sum_{\nu < \mu} 2^{\aleph_\nu}$ . Thus  $S$  is of cardinal  $\aleph_\nu$ . If  $\aleph_{\mu+1} = 2^{\aleph_\mu}$ , then  $\aleph_\nu = \aleph_\mu$  and  $S$  is of cardinal  $\aleph_\mu$ . As we have seen,  $I_\mu$  is of cardinal  $\aleph_\mu$ . We may say, therefore, that our knowledge of whether the cardinal of  $S$  is as small as that of  $I_\mu$  depends upon the validity of the generalized continuum hypothesis.

**11. The order types  $\alpha$ -dense and  $\alpha$ -scattered.** Starting with the order property "has cardinal less than  $\aleph_\lambda$ ," we were led to the order properties  $\aleph_\lambda$ -dense and  $\aleph_\lambda$ -scattered. In a similar fashion, starting instead with an arbitrary order property, we are led to the properties we now describe. An order every proper segment of which has more than two elements and contains a suborder of property  $\alpha$  will be termed  $\alpha$ -dense. An order containing no  $\alpha$ -dense

suborders will be termed  $\alpha$ -scattered. If  $\alpha$  is the property of containing  $\aleph_\lambda$  or more elements,  $\alpha$ -dense and  $\alpha$ -scattered are equivalent to the previously defined properties  $\aleph_\lambda$ -dense and  $\aleph_\lambda$ -scattered, respectively. Substituting  $\alpha$ -dense for dense and  $\alpha$ -scattered for scattered in the proof of Theorem 5, we may prove

**THEOREM 16.** *If  $\alpha$  is any order property,  $\alpha$ -scattered is transitive.*

We note the following alternative wording of the definition of  $\alpha$ -scattered. An order  $A$  is  $\alpha$ -scattered if every suborder is not  $\alpha$ -dense. I.e., there exists a proper segment of every suborder which either has exactly two elements or is such that every suborder has property  $\alpha^N$ , where  $\alpha^N$  denotes the order property: "not of type  $\alpha$ ." In particular, an  $\aleph_0$ -scattered order, that is, a scattered order, may be defined as an order which has the property that every suborder of it contains a segment consisting of exactly two elements. By comparing these definitions of scattered and  $\alpha$ -scattered, it becomes apparent the type  $\alpha$ -scattered includes scattered orders as subtype. This would, of course, also be inferred by Theorems 7 and 16.

Since  $\alpha$ -scattered is transitive, it furnishes a second decomposition of every order as follows. Every order is either  $\alpha$ -scattered or the sum of  $\alpha$ -scattered orders over an order  $B$  no proper segment of which has property  $\alpha$ -scattered. That is to say, every proper segment of  $B$ , if it exists, contains an  $\alpha$ -dense order. We conclude that

**THEOREM 17.** *If  $\alpha$  is any order property and  $A$  an order,  $A$  is either  $\alpha$ -scattered or the sum of  $\alpha$ -scattered orders over an  $\alpha$ -dense order.*

$\alpha$ -scattered, being transitive, is equivalent to some property  $\beta^T$  when  $\beta$  is properly chosen. We may ask for an "economical" way to describe  $\beta$  in terms of  $\alpha$ . We have, of course, an "extravagant" solution if we set  $\beta$  equal to  $\alpha$ -scattered. A more "thrifty" answer is developed as follows. We have seen that every suborder of an  $\alpha$ -scattered order has a proper segment which either consists of exactly two elements or is such that every suborder has property  $\alpha^N$ . Thus  $\alpha$ -scattered orders contain scattered orders but no  $\alpha$ -orders. We try

**THEOREM 18a.** *If  $\alpha$  is any order property, the order property  $\alpha$ -scattered is equivalent to the order property  $\bar{\alpha}^T$ , where  $\bar{\alpha}$  is the property of containing no suborders of property  $\alpha$ .*

**Proof.** An  $\bar{\alpha}$ -order is  $\alpha$ -scattered since it contains no  $\alpha$ -orders and hence, surely, no  $\alpha$ -dense order. Therefore,  $\alpha$ -scattered is a transitive order type which includes  $\bar{\alpha}$ -orders as subtype. Since  $\bar{\alpha}^T$  is the minimal transitive order type which includes  $\bar{\alpha}$  as subtype,  $\alpha$ -scattered includes  $\bar{\alpha}^T$  as subtype. Suppose, now,  $A$  is an  $\alpha$ -scattered order. By Theorem 11, either  $A$  has property  $\bar{\alpha}^T$  or is the sum of  $\bar{\alpha}^T$ -orders over an order  $B$  no proper segment of which has property  $\bar{\alpha}^T$ . In the latter case  $B$  has, in particular, no proper segment with

property  $\bar{\alpha}$ . Consequently, if  $B$  exists, every proper segment of  $B$  contains a suborder with property  $\alpha$ —i.e., contains an  $\alpha$ -dense suborder. Thus  $A$  would contain an  $\alpha$ -dense suborder, contrary to hypothesis. We conclude that  $A$  has property  $\bar{\alpha}^T$ . The latter property is therefore equivalent to the property  $\alpha$ -scattered.

We determine a construction for  $\alpha$ -scattered in terms of operations on  $\alpha$ -orders. Let us consider  $\bar{\alpha}$ , the order property "containing no  $\alpha$ -order as suborder." This is equivalent to the property "every suborder has property  $\alpha^N$ ." Now, if we denote by  $\alpha^R$  the property of being a superorder of an  $\alpha$ -order, the property  $\bar{\alpha}$  is equivalent to the property  $\alpha^{RN}$ . Accordingly, Theorem 18a may be written:

THEOREM 18b.  $\alpha$ -scattered  $= \alpha^{RNDEI}$ .

Just as  $\alpha^{DD} = \alpha^D$ , so  $\alpha^{RR} = \alpha^R$ . We shall term  $\alpha^R$  *rising*, and, in general, if for an order property  $\alpha$ ,  $\alpha^R$  is equivalent to  $\alpha$ , that is, if every superorder of an  $\alpha$ -order is an  $\alpha$ -order, we shall term  $\alpha$  *rising*. It is clear  $\alpha^R$  is the minimal rising property implied by  $\alpha$ . On the other hand, the type  $\alpha^{RN}$  has the property that a suborder of an  $\alpha^{RN}$ -order is an  $\alpha^{RN}$ -order, and we term  $\alpha^{RN}$  *falling*. More generally, we shall name an order property  $\alpha$  *falling* if every suborder of an  $\alpha$ -order is an  $\alpha$ -order. If  $\alpha$  is any order type we shall denote by  $\alpha^F$  the property of being a suborder of an  $\alpha$ -order. Manifestly  $\alpha^F$  is the minimal falling supertype of  $\alpha$ . Again we have a "closure equation"  $\alpha^{FF} = \alpha^F$ . We note also, for future reference, that if  $\alpha$  is a falling property,  $\alpha^N$  is a rising property and conversely. Thus  $\alpha^{FN}$  is rising and  $\alpha^{RN}$  is falling, for all  $\alpha$ .

In the equation  $\alpha$ -scattered  $= \alpha^{RNDEI}$ , we may regard the  $RNDEI$ -operation as a "solution for  $X$ " in the conditional equation  $\alpha$ -scattered  $= \alpha^X$ . Conversely, we inquire as to possible "solutions" of the equation  $\alpha^X$ -scattered  $= \alpha$ , where by  $X$  we have in mind an operation corresponding to some combination of the letters  $F, R, I$ , etc. Let us assume there exists a solution.  $\alpha^X$ -scattered is a transitive order type. Clearly, it is also falling. Therefore  $\alpha^F = \alpha$ ,  $\alpha^T = \alpha$ , and  $\alpha^T = (\alpha^F)^T = \alpha^{FT}$ . Combining, we obtain  $\alpha^X$ -scattered  $= \alpha^{FT}$ . But, by Theorem 18b, we may substitute  $\alpha^{XRNDI} = \alpha^{XRNT}$  for  $\alpha^X$ -scattered. Our conditional equation becomes  $\alpha^{XRNT} = \alpha^{FT}$ . This is true if  $\alpha^{XRN} = \alpha^F$ , or  $\alpha^{XR} = \alpha^{FN}$ . The latter equation is equivalent to  $\alpha^{XR} = \alpha^{FNR}$ , since  $\alpha^{FN}$  is rising. Finally,  $\alpha^{XR} = \alpha^{FNR}$  is implied by  $\alpha^X = \alpha^{FN}$ . We state

THEOREM 19. If  $\alpha$  is any order type, the minimal falling and transitive supertype of  $\alpha$  is  $\alpha^{FT} = \alpha^{FN}$ -scattered.

**Proof.** We prove the equivalence of  $\alpha^{FT}$  and  $\alpha^{FN}$ -scattered directly. For, since  $\alpha^{FNR} = \alpha^{FN}$ ,  $\alpha^{FN}$ -scattered  $= (\alpha^{FN})^{RNT} = \alpha^{FNRNT} = \alpha^{FNNT} = \alpha^{FT}$ .  $\alpha^{FT}$  is minimal for, by Theorem 10,  $\alpha^{FT} = (\sigma\alpha^F)^I = (\sigma\alpha^F)^I$  where  $\sigma$  is the order type scattered. The product of two falling order types is, in every case, falling. Thus  $\sigma\alpha^F$  is falling. In Theorem 3 we proved that  $\alpha^I$  is descending if  $\alpha$  is



descending. In a similar fashion we may show  $\alpha^I$  is falling if  $\alpha$  is falling. We conclude  $(\sigma\alpha^F)^I = \alpha^{FT}$  is falling. Furthermore,  $\alpha^{FT}$  is transitive since the  $T$ -operation is performed last. Every falling and transitive order type  $\beta$  which includes  $\alpha$  includes  $\alpha^{FT}$ . For  $\beta$  includes  $\alpha$  implies  $\beta$  includes the minimal  $\alpha^F$ . Since  $\beta$  is transitive, includes  $\alpha^F$ , and  $\alpha^{FT}$  is the minimal transitive type which includes  $\alpha^F$ ,  $\beta$  includes  $\alpha^{FT}$ . Thus  $\alpha^{FT} = \alpha^{FN}$ -scattered satisfies the requisite minimal condition of the theorem.

We are now in a position to "solve" for  $X$  in the equation  $\alpha^X$ -scattered  $= \alpha$ . There is a solution if and only if  $\alpha = \alpha^F = \alpha^{FT}$ . Then  $\alpha^{FN}$ -scattered  $= \alpha^{FT} = \alpha$ ; or  $\alpha^N$ -scattered  $= \alpha$ . Thus  $X = N$  is a solution.

We next establish the following characterization for the type  $\alpha$ -scattered.

**THEOREM 20.** *If  $\alpha$  is any order type, the order type  $\alpha$ -scattered is the minimal falling and transitive supertype of  $\alpha^{RN}$ .*

**Proof.** The minimal falling and transitive type which includes  $\alpha^{RN}$  is, by Theorem 19,  $\alpha^{RNFT}$ . Since  $\alpha^{RN}$  is falling,  $\alpha^{RNF} = \alpha^{RN}$  and  $\alpha^{RNFT} = \alpha^{RNT}$ . But  $\alpha^{RNT} = \alpha$ -scattered. Thus  $\alpha^{RNFT} = \alpha$ -scattered, proving the theorem.

The above results show the class of order properties  $\alpha$ -scattered is identical with the class of order properties which are both transitive and falling. Moreover, a transitive order property is not, in every case, a falling order property. For example,  $\alpha^T$ , where  $\alpha$  is the order type of the continuum, does not include the order type of the rational numbers as subtype.

In regard to the notion  $\alpha$ -dense, it turns out the equation  $\alpha^X = (\alpha\text{-dense})^R$  has a "solution" <sup>(13)</sup>. For an order is  $\alpha$ -scattered if no  $\alpha$ -dense suborders exist. An equivalent statement is that  $\alpha$ -scattered  $= (\sigma\text{-dense})^{RN}$ . Thus  $(\alpha\text{-scattered})^N = \alpha^{RNTN} = (\alpha\text{-dense})^R$ .

**12. Properties of orders with  $\aleph_\lambda$  elements.** The decomposition of Theorem 4a shows that

**THEOREM 21.** *Every order of cardinal  $\aleph_\lambda$  containing neither  $\omega_\lambda$  nor  $\omega_\lambda$ -reversed, with  $\aleph_\lambda$  regular, is the sum of orders each of cardinal less than  $\aleph_\lambda$  over an  $\aleph_\lambda$ -dense order.*

**Proof.** The property  $\alpha$  of containing less than  $\aleph_\lambda$  elements is iterative and descending. With this choice of  $\alpha$ , Theorem 4a becomes the above theorem.

Thus every order of cardinal  $\aleph_\lambda$ , containing neither  $\omega_\lambda$  nor  $\omega_\lambda$ -reversed, with  $\aleph_\lambda$  regular, contains an  $\aleph_\lambda$ -dense order. It follows that

<sup>(13)</sup> It seems that in the equation  $\alpha^X = \alpha\text{-dense}$  there is no solution for  $X$  in terms of the letters  $D, E, I, F, R, N$ . We may, however, construct an  $\alpha$ -dense order, once  $\alpha$  is given, as follows: Let  $A_1, A_2, \dots, A_n, \dots$  be a series of  $\alpha$ -orders. We may form a development  $a_1a_2 \dots a_n \dots$ , where the entry  $a_n$  is an element of  $A_n$ . The set of all such developments, ordered lexicographically, constitutes an  $\alpha$ -dense order. The set of all such orders determines an order type which might also appropriately be termed an  $\omega$ th power of  $\alpha$ . Cf. §3.

**THEOREM 22.** *Every order of regular cardinal  $\aleph_\lambda$  contains either  $\omega_\lambda$ , or  $\omega_\lambda$ -reversed, or an  $\aleph_\lambda$ -dense order.*

Order types  $\alpha$  which may serve as a starting point in the formation of new order types and in the study of the structure of orders are: dense, closed, perfect, of cardinal  $\aleph_\lambda$ ,  $\omega_\lambda$ ,  $\rho_\lambda$ <sup>(14)</sup>, etc. We may then form the order types  $\alpha^D$ ,  $\alpha^R$ ,  $\alpha^I$ ,  $\alpha^E$ ,  $\alpha^F$ ,  $\alpha^N$  and any combination of these to form new order types  $\beta$ . With each of these  $\beta$ 's is associated the transitive types  $\beta^T$  and  $\beta$ -scattered each of which provides a segmental decomposition of every order. A number of these associated types and decompositions have been considered in this paper. Others, of possible interest, we leave to future investigation. We note, too, the possibility of introducing order types associated with  $\alpha$  by other means such as classification according to properties of Dedekind cuts, properties of initial segments, etc.<sup>(15)</sup>.

<sup>(14)</sup> See Hausdorff, *Mengenlehre*, pp. 180-185.

<sup>(15)</sup> See Hausdorff, *Mengenlehre*, pp. 142-147.

ST. MICHAEL'S COLLEGE,  
WINOOSKI PARK, VT.

## EXPANSIONS OF ANALYTIC FUNCTIONS

BY

R. P. BOAS, JR.

**Introduction.** There is an extensive literature dealing with the problem of expanding analytic functions of a complex variable in generalized Taylor series of the form

$$(1) \quad f(z) = \sum_{n=0}^{\infty} c_n g_n(z),$$

where the  $g_n(z)$  are, in a suitable sense, "nearly" the functions  $z^n$ <sup>(1)</sup>. If  $g_n(z) = z^n [1 + h_n(z)]$ , where the  $h_n(z)$  are analytic and bounded in a circle  $|z| < r$  and vanish at  $z=0$ , and  $f(z)$  is analytic in  $|z| < r$ , the possibility of an expansion of the form (1) was established by S. Pincherle [9]; the series converges to  $f(z)$  in some circle  $|z| < s$ , where in general  $s < r$ . Much of the later work has been devoted to obtaining better estimates for the number  $s$ . In this paper, a new attack on the problem is developed; it eliminates rearrangements of power series, and uses a criterion for "nearness" of two sequences of functions which is essentially contained in work of Paley and Wiener [26, p. 100] (where it is applied to another problem). The results include some of those of G. S. Ketchum [4], which are the most precise yet obtained, and in part go beyond them. Well known expansion theorems of G. D. Birkhoff [1] and J. L. Walsh [17] are also obtained.

The simplest of my results (and the most convenient one for applications) is that if the functions  $g_n(z)$  in (1) are of the form specified above, and if the  $h_n(z)$  have a common majorant  $h(z)$  for large  $n$  (that is, if the coefficients in the power series of  $h_n(z)$  are less in absolute value than the corresponding coefficients of  $h(z)$ ), then the expansion (1) converges to  $f(z)$  in  $|z| < s$  if  $h(s) < 1$ . For example, if  $1 + h_n(z) = e^{\alpha_n z}$ , with  $\limsup_{n \rightarrow \infty} |\alpha_n| \leq 1$ , we may take  $h(z) = e^{(1+\epsilon)z} - 1$  (with any positive  $\epsilon$ ), so that the region of convergence of (1) is at least  $|z| < \log 2$ ; I have not been able to establish convergence in a larger region than  $|z| < 1/e$  by using the theorems in the literature<sup>(2)</sup>.

It is also possible to restrict linear combinations of the coefficients of the

---

Presented to the Society, April 27, 1940; received by the editors March 18, 1940.

(<sup>1</sup>) The bibliography at the end of this paper contains all the references which I have found (without however making an intensive search of the literature) on general expansions of this type. For special theorems, other than those considered in this paper, see especially G. S. Ketchum [4]. (Numbers in brackets refer to the bibliography.)

(<sup>2</sup>) Added in proof: Ibragimoff [32] has proved that every function analytic in  $|z| < s$  is the uniform limit in  $|z| \leq s' < s$  of a sequence of linear combinations of the functions in question if  $s \leq \log 2$ .

$h_n(z)$  instead of the coefficients themselves; this can be done by a method different from that used by G. S. Ketchum in obtaining the first such results (see §5). Another generalization consists in modifying the assumption that the functions  $g_n(z)$  should have precisely the form  $z^n[1+h_n(z)]$  (see Theorem 6.4).

The expansion theorems of this paper were originally developed in the hope (which has so far proved illusory) of settling a conjecture concerning the values taken by derivatives of entire functions. However, I have obtained some new results in this field. In particular, I prove the following theorem (Theorem 7.1): If  $f(z)$  is an entire function of exponential type  $k < \log 2$ , with  $f(0) = 1$ , and if the points  $\alpha_n$  ( $n = 0, 1, 2, \dots$ ) are in the circle  $|z| \leq 1$ , then for every  $r < k$

$$\sum_{n=0}^{\infty} \frac{|f^{(n)}(\alpha_n)|^2}{r^{2n}} \geq 2e^r - e^{2r}.$$

This generalizes a theorem of S. Takenaka<sup>(\*)</sup> which states that  $f^{(n)}(\alpha_n)$  cannot be zero for all  $n$ .

Many of the papers listed in the bibliography treat, besides the convergence of the series (1), the existence of systems of functions biorthogonal to the  $g_n(z)$ , the form of the coefficients in (1), etc. These problems are not considered in this paper, although its methods could be made to furnish information about them.

Some of the results of this paper were announced, with indications of the proofs, in a note in the Proceedings of the National Academy of Sciences<sup>(\*)</sup>.

**1. Abstract expansion theorems.** We consider a normed complex linear space  $E$ , and a sequence  $G = \{x_n\}$  of elements of  $E$ .  $G$  is said to be a fundamental set if the set of all finite linear combinations of elements of  $G$  is everywhere dense in  $E$ ; that is, if for every  $y \in E$  there exist complex numbers  $c_{k,n}$  such that

$$(1.1) \quad y = \lim_{n \rightarrow \infty} \sum_{k=1}^n c_{k,n} x_k.$$

$G$  is said to be a base if every element  $y \in E$  has a unique representation as an infinite series of multiples of elements of  $E$ ; that is, if for every  $y \in E$  there exists a unique sequence of complex numbers  $c_k$  such that

$$(1.2) \quad y = \lim_{n \rightarrow \infty} \sum_{k=1}^n c_k x_k.$$

The following theorem states in effect that a sequence sufficiently near

<sup>(\*)</sup> See J. M. Whittaker [30, p. 44]; Takenaka [29].

<sup>(\*)</sup> Vol. 26 (1940), pp. 139-143.

another sequence which is a fundamental sequence or a base is also a fundamental sequence or a base.

**THEOREM 1.1.** *Let the sequences  $G = \{x_n\}$  and  $H = \{y_n\}$  have the property that for some number  $\lambda$  ( $0 < \lambda < 1$ ), and for all finite sequences  $a_1, a_2, \dots, a_N$  of complex numbers,*

$$(1.3) \quad \left\| \sum_{n=1}^N a_n(x_n - y_n) \right\| \leq \lambda \left\| \sum_{n=1}^N a_n x_n \right\|.$$

Then

- (i) if  $G$  is a fundamental set, so is  $H$ ;
- (ii) if  $E$  is complete and  $G$  is a base,  $H$  is a base.

In case (ii), furthermore, if the element  $x \in E$  has the expansion

$$\sum_{k=1}^{\infty} c_k y_k,$$

the coefficients  $c_k$  have the property

$$(1.4) \quad \left\| \sum_{k=1}^{\infty} c_k x_k \right\| \leq \frac{1}{1-\lambda} \|x\|.$$

Theorem 1.1 (ii), in the special form which it assumes when  $G$  is a normal orthogonal base, was given (with a proof which applies to a general base  $G$ ) by Paley and Wiener [26, p. 100] for the Hilbert space  $L^2(-\pi, \pi)$ . For a general Banach space, the proof given by Paley and Wiener needs only formal modifications; in this paper, Theorem 1.1 (ii) will be used almost exclusively for Hilbert spaces, and is consequently established by the proof of Paley and Wiener (since all realizations of abstract Hilbert space are equivalent). We omit the proof of Theorem 1.1 (ii).

The proof of Theorem 1.1 (i) is considerably simpler; this part would be sufficient for the applications which will be made in §7 to derivatives of analytic functions. We suppose that  $G$  is fundamental, that  $H$  is not, and that (1.3) is satisfied. Then there is a linear<sup>(\*)</sup> functional  $f$ , defined on  $E$ , such that  $f(y_n) = 0$ ,  $n = 1, 2, \dots$ , while  $f(z) \neq 0$  for some  $z$ . Let

$$f(x_n) = f(x_n - y_n) = c_n \quad (n = 1, 2, \dots).$$

Let  $M = \|f\|$ ; that is, let  $M$  be the smallest number such that, for all  $x \in E$ ,  $|f(x)| \leq M\|x\|$ . Then for any sequence  $\{a_n\}$

$$\left| \sum_{n=1}^N a_n c_n \right| \leq M \left\| \sum_{n=1}^N a_n (x_n - y_n) \right\| \leq M\lambda \left\| \sum_{n=1}^N a_n x_n \right\|.$$

(\*) "Linear" means "distributive and continuous," as in Banach's book [21].



Hence<sup>(6)</sup> there is a linear functional  $g$ , defined on  $E$ , such that  $g(x_n) = c_n$  ( $n = 1, 2, \dots$ ), and  $\|g\| \leq M\lambda < M$ . But since  $\{x_n\}$  is a fundamental set and  $f(x_n) - g(x_n) = 0$  ( $n = 1, 2, \dots$ ), we must have  $f(x) = g(x)$  for every  $x$ , and consequently  $M = \|f\| = \|g\| \leq \lambda M < M$ , a contradiction; for  $M$  is not zero because  $f(z) \neq 0$  for some  $z$ .

**2. General expansions of analytic functions.** We now apply Theorem 1.1 to the spaces  $H_p(r)$  whose elements are functions  $f(z)$  analytic in  $|z| < r$ , belonging to  $L^p$  ( $p \geq 1$ ) in this circle; that is, each function  $f(z)$  is assumed to satisfy<sup>(7)</sup>

$$(2.1) \quad \left\{ \frac{1}{2\pi} \int_0^{2\pi} |f(\rho e^{i\theta})|^p d\theta \right\}^{1/p} \leq A, \quad 0 \leq \rho < r,$$

where  $A$  depends only on  $f$ . It is well known<sup>(8)</sup> that if  $f(z)$  satisfies (2.1) it has boundary values almost everywhere on  $|z| = r$ , and that the boundary function belongs to  $L^p$ . We complete the definition of  $H_p(r)$  by defining the norm of  $f(z)$  by the relation

$$\|f\| = \left\{ \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right\}^{1/p}.$$

We introduce, to save repetition, the following

**DEFINITION.** A sequence  $\{f_n(z)\}$  of functions analytic in  $|z| < r$  and belonging to some class  $H_p(r)$  ( $1 \leq p \leq \infty$ ) has *Property T* in  $|z| < r$  if every function  $f(z)$  analytic in  $|z| < r$  and continuous in  $|z| \leq r$  can be expanded in a unique series of the form

$$(2.2) \quad f(z) = \sum_{n=1}^{\infty} c_n f_n(z),$$

the series converging uniformly in every circle  $|z| \leq r' < r$ . If furthermore the series in (2.2) converges uniformly in  $|z| \leq r$ , the sequence has *Property T<sub>∞</sub>*.

The sequence  $(1, z, z^2, \dots)$  is an obvious example of a sequence having *Property T<sub>∞</sub>* in any circle.

Applied to the spaces  $H_p(r)$ , Theorem 1.1 yields

**THEOREM 2.1.** Let  $\{f_n(z)\}$  and  $\{g_n(z)\}$  be two sequences of elements of  $H_p(r)$ , such that for some numbers  $p$  and  $\lambda$  ( $1 \leq p \leq \infty$ ,  $0 < \lambda < 1$ ), and for all sets of complex numbers  $a_1, a_2, \dots, a_N$

<sup>(6)</sup> Banach [21, p. 56]. The result remains valid for complex linear spaces: see Bohnenblust and Sobczyk [23].

<sup>(7)</sup> Expressions involving  $p$  are to be interpreted according to the usual conventions when  $p = \infty$ : that is, as the limits as  $p \rightarrow \infty$  of the corresponding expressions for finite  $p$ .

<sup>(8)</sup> See, e.g., Zygmund [31, p. 162].

$$(2.3) \quad \left\{ \int_0^{2\pi} \left| \sum_{n=1}^N a_n [f_n(re^{i\theta}) - g_n(re^{i\theta})] \right|^p d\theta \right\}^{1/p} \\ \leq \lambda \left\{ \int_0^{2\pi} \left| \sum_{n=1}^N a_n f_n(re^{i\theta}) \right|^p d\theta \right\}^{1/p}.$$

Then, in  $|z| < r$ ,  $\{g_n\}$  has Property T if  $\{f_n\}$  has Property T; if (2.3) is satisfied with  $p = \infty$ ,  $\{g_n\}$  has Property  $T_\infty$  if  $\{f_n\}$  has Property  $T_\infty$ . Moreover, if the expansion of  $f(z)$  in terms of  $\{g_n(z)\}$  has the form

$$(2.4) \quad f(z) = \sum_{n=1}^{\infty} c_n g_n(z),$$

the coefficients  $c_n$  have the property

$$(2.5) \quad \left\{ \int_0^{2\pi} \left| \sum_{k=1}^{\infty} c_k f_k(re^{i\theta}) \right|^p d\theta \right\}^{1/p} \leq \frac{1}{1-\lambda} \left\{ \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right\}^{1/p}.$$

The direct deduction from Theorem 1.1 is that the series in (2.4) converges to  $f(z)$  in the topology of  $H_p(r)$ . In case  $p = \infty$ , this is the desired conclusion. Otherwise, if  $|z| \leq s < r$  we have

$$\left| f(z) - \sum_{n=1}^N c_n g_n(z) \right| = \left| \frac{1}{2\pi i} \int_{|w|=r} \left\{ f(w) - \sum_{n=1}^N c_n g_n(w) \right\} \frac{dw}{w-z} \right|;$$

an application of Hölder's inequality shows that

$$\lim_{N \rightarrow \infty} \left| f(z) - \sum_{n=1}^N c_n g_n(z) \right| = 0,$$

uniformly in  $|z| \leq s$ .

We shall use Theorem 2.1 most frequently in the special case when  $f_n(z) = z^{n-1}$ . It then becomes

**THEOREM 2.2.** *The sequence  $\{g_n(z)\}$  has Property T in  $|z| < r$  if, for all sets of complex numbers  $a_0, a_1, \dots, a_N$ ,*

$$(2.6) \quad \left\{ \int_0^{2\pi} \left| \sum_{n=0}^N a_n [r^n e^{in\theta} - g_n(re^{i\theta})] \right|^p d\theta \right\}^{1/p} \\ \leq \lambda \left\{ \int_0^{2\pi} \left| \sum_{n=0}^N a_n r^n e^{in\theta} \right|^p d\theta \right\}^{1/p},$$

where  $p$  and  $\lambda$  satisfy  $1 \leq p \leq \infty$ ,  $0 < \lambda < 1$ . If (2.6) is true with  $p = \infty$ , the sequence has Property  $T_\infty$ .

From Theorem 2.2 we can deduce in a few lines the following generalization of expansion theorems of G. D. Birkhoff [1] and J. L. Walsh [17].

**THEOREM 2.3.** *If the functions  $g_n(z)$  are analytic in  $|z| < r$ , continuous in  $|z| \leq r$ , and satisfy*

$$(2.7) \quad \sum_{n=0}^{\infty} r^{-2n} |g_n(z) - z^n|^2 < 1, \quad |z| = r,$$

*the series converging uniformly, then the set  $\{g_n(z)\}$  has Property  $T_{\infty}$  in every circle  $|z| < s \leq r$ .*

In Birkhoff's theorem, (2.7) is replaced by

$$(2.8) \quad \sum_{n=0}^{\infty} r^{-n} |g_n(z) - z^n| < 1, \quad |z| = r;$$

this condition implies (2.7), by Cauchy's inequality. In Walsh's theorem (2.7) holds, and in addition the series in (2.8) is assumed to converge.

We apply the case  $p = \infty$  of Theorem 2.2. The sum of the series in (2.7) is continuous on  $|z| = s$ , when  $s \leq r$ , and so has a maximum  $\lambda^2 < 1$ . We have, with  $z = e^{i\theta}$ ,

$$\begin{aligned} \max_{0 \leq \theta \leq 2\pi} \left| \sum_{n=0}^N a_n [g_n(z) - z^n] \right| &\leq \left( \sum_{n=0}^N |a_n|^2 s^{2n} \right)^{1/2} \max_{0 \leq \theta \leq 2\pi} \left( \sum_{n=0}^N s^{-2n} |g_n(z) - z^n|^2 \right)^{1/2} \\ &\leq \lambda \left( \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^N a_n z^n \right|^2 d\theta \right)^{1/2} \\ &\leq \lambda \max_{0 \leq \theta \leq 2\pi} \left| \sum_{n=0}^N a_n z^n \right|. \end{aligned}$$

This establishes (2.6) with  $p = \infty$ , and Theorem 2.3 follows.

In this section we have applied part (ii) of Theorem 1.1. The weaker part (i) would yield a weak form of Property T with the uniformly convergent series replaced by a uniformly convergent sequence of linear combinations.

**3. Criteria for the existence of expansion theorems.** From now on, we shall use Theorem 2.2 exclusively in the case  $p = 2$ , which is the case in which criteria for the validity of (2.6) are most easily set up. Our functions  $g_n(z)$  will, in this section, be of the form

$$(3.1) \quad g_n(z) = z^n [1 + h_n(z)] \quad (n = 0, 1, 2, \dots),$$

where

$$(3.2) \quad h_n(z) = \sum_{k=1}^{\infty} \gamma_k^{(n)} z^k \quad (|z| < r_0).$$

We assume to begin with that the  $h_n(z)$  have a common majorant  $h(z)$ ; that is, that

$$|\gamma_k^{(n)}| \leq \delta_k \quad (k = 1, 2, \dots; n = 0, 1, 2, \dots),$$

where  $h(z) = \sum_{k=1}^{\infty} \delta_k z^k$  ( $|z| < r_0$ ). This restriction will be considerably relaxed in §4. We introduce the quantity  $K_p$  by the definition

$$(3.3) \quad K_p^2 = \frac{1}{2\pi} \int_0^{2\pi} |h(\rho e^{i\theta})|^2 d\theta = \sum_{k=1}^{\infty} \delta_k^2 \rho^{2k} \quad (\rho < r_0).$$

**THEOREM 3.1.** *The functions  $g_n(z)$  have Property T in any circle  $|z| < s$  provided that one of the following conditions is satisfied:*

$$(3.4) \quad h(s) < 1,$$

$$(3.5) \quad s < \sup_{0 \leq \rho < r_0} \frac{\rho}{(K_p^2 + 1)^{1/2}}.$$

We have to verify (2.6) with  $p = 2, r = s$ , for an arbitrary set  $(a_0, a_1, \dots, a_N)$ . We write

$$a'_n = \begin{cases} a_n, & n = 0, 1, \dots, N, \\ 0, & n > N; \end{cases}$$

$$\psi(z) = \sum_{n=0}^{\infty} |a'_n| z^n, \quad \psi_m(z) = \sum_{n=0}^m |a'_n| z^n.$$

Then condition (2.6) takes the form

$$(3.6) \quad \begin{aligned} \Phi(r) &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^{\infty} a'_n r^n e^{in\theta} h_n(re^{i\theta}) \right|^2 d\theta \\ &\leq \lambda^2 \sum_{n=0}^{\infty} |a'_n|^2 r^{2n}. \end{aligned}$$

$\Phi(r)$  can be rewritten as follows.

$$(3.7) \quad \begin{aligned} \Phi(r) &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^{\infty} a'_n r^n e^{in\theta} \sum_{k=1}^{\infty} \gamma_k^{(n)} r^k e^{ik\theta} \right|^2 d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{m=1}^{\infty} r^m e^{im\theta} \sum_{n=0}^{m-1} a'_n \gamma_{m-n}^{(n)} \right|^2 d\theta \\ &= \sum_{m=1}^{\infty} r^{2m} \left| \sum_{n=0}^{m-1} a'_n \gamma_{m-n}^{(n)} \right|^2. \end{aligned}$$

In the first place, we evidently have

$$(3.8) \quad \Phi(r) \leq \sum_{m=1}^{\infty} r^{2m} \left\{ \sum_{n=0}^{m-1} |a'_n| \delta_{m-n} \right\}^2.$$

If we retrace the steps in (3.7), we then find

$$\begin{aligned} \Phi(r) &\leq \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^{\infty} a'_n | r^n e^{in\theta} \sum_{k=1}^{\infty} \delta_k r^k e^{ik\theta} \right|^2 d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} |h(re^{i\theta}) \psi(re^{i\theta})|^2 d\theta \\ &\leq \{h(r)\}^2 \frac{1}{2\pi} \int_0^{2\pi} |\psi(re^{i\theta})|^2 d\theta \\ &= \{h(r)\}^2 \sum_{n=0}^{\infty} |a'_n|^2 r^{2n}. \end{aligned}$$

Thus if  $h(r) < 1$ , (3.6) is satisfied; this proves Theorem 3.1 under condition (3.4).

We now observe that the expression

$$\sum_{n=0}^{m-1} |a'_n| \delta_{m-n}$$

which occurs on the right of (3.8) is the coefficient of  $z^m$  in the power series of  $\psi_{m-1}(z)h(z)$ , and consequently can be written as

$$\frac{1}{2\pi i} \int_{|z|=\rho} \frac{h(z)\psi_{m-1}(z)}{z^{m+1}} dz \quad (|z| = \rho < r_0).$$

Hence its square does not exceed

$$\frac{1}{\rho^{2m}} \frac{1}{2\pi} \int_0^{2\pi} |h(\rho e^{i\theta})|^2 d\theta \frac{1}{2\pi} \int_0^{2\pi} |\psi_{m-1}(\rho e^{i\theta})|^2 d\theta = \frac{K_\rho^2}{\rho^{2m}} \sum_{n=0}^{m-1} |a'_n|^2 \rho^{2n}.$$

From (3.8) we now obtain, if  $r < \rho$ ,

$$\begin{aligned} \Phi(r) &\leq K_\rho^2 \sum_{m=1}^{\infty} \left(\frac{r}{\rho}\right)^{2m} \sum_{n=0}^{m-1} |a'_n|^2 \rho^{2n} \\ &= K_\rho^2 \sum_{n=0}^{\infty} |a'_n|^2 r^{2n} \sum_{m=1}^{\infty} \left(\frac{r}{\rho}\right)^{2m} = K_\rho^2 \frac{r^2}{\rho^2 - r^2} \sum_{n=0}^{\infty} |a'_n|^2 r^{2n}. \end{aligned}$$

Then (3.6) is satisfied if we can choose  $\rho$  so that  $K_\rho^2 r^2 / (\rho^2 - r^2) < 1$ , or so that

$$r^2 < \frac{\rho^2}{K_\rho^2 + 1}.$$



This shows that the relation (3.6) is satisfied if  $r=s$  and  $s$  satisfies inequality (3.5).

In earlier theorems of the same type as Theorem 3.1, the conditions have restricted the coefficients  $\delta_k$  of the function  $h(z)$  majorizing the  $h_n(z)$ ; here we restrict only the behavior of  $h(z)$  in the large<sup>(9)</sup>. Two less precise known theorems can be deduced as corollaries of Theorem 3.1.

**THEOREM 3.2<sup>(10)</sup>.** *If  $|h_n(z)| \leq M(\rho)$  ( $n=0, 1, 2, \dots$ ;  $|z| \leq \rho$ ), then the functions  $g_n(z)$  have Property T in  $|z| < s$  if*

$$(3.9) \quad s < \sup_{0 \leq \rho < r_0} \frac{\rho}{M(\rho) + 1}.$$

In fact, Cauchy's inequalities for derivatives yield, for  $0 < \rho < r_0$ ,

$$|\gamma_k^{(n)}| \leq M(\rho)\rho^{-k} \quad (n=0, 1, 2, \dots; k=1, 2, \dots).$$

Consequently, if  $\rho < r_0$ , we can take  $\delta_k = M(\rho)\rho^{-k}$  ( $k=1, 2, \dots$ ); we then have  $h(r) = rM(\rho)/(\rho-r)$ , and  $h(r) < 1$  if  $r < \rho/[M(\rho)+1]$ . If we choose  $\rho$  in the most favorable way, Theorem 3.2 follows from Theorem 3.1.

**THEOREM 3.3<sup>(11)</sup>.** *If*

$$(3.10) \quad L_\rho = \sup_{1 \leq k < \infty} \delta_k \rho^k \quad (\rho < r_0),$$

*the functions  $g_n(z)$  have Property T in  $|z| < s$  if*

$$(3.11) \quad s < \sup_{0 \leq \rho < r_0} \frac{\rho}{L_\rho + 1}.$$

We have

$$\begin{aligned} \delta_k &\leq \rho^{-k} L_\rho; \\ h(r) &\leq L_\rho \sum_{n=1}^{\infty} \left(\frac{r}{\rho}\right)^n \quad (r < \rho) \\ &= L_\rho \frac{r}{\rho - r}; \end{aligned}$$

hence  $h(s) < 1$  if  $s$  satisfies (3.11). The conclusion follows from Theorem 3.1.

For use in §5, we note the following property of the coefficients in the expansion whose existence is established by Theorems 3.1 and 3.2. If  $f(z)$  has the expansion

<sup>(9)</sup> However, Theorem 3.1 (even as generalized in §4) does not seem to include Theorem III of G. S. Ketchum [4].

<sup>(10)</sup> Narumi [7], Takenaka [15], G. S. Ketchum [4].

<sup>(11)</sup> Graesser [2]. See also G. S. Ketchum [4, p. 215, footnote].

$$f(z) = \sum_{n=0}^{\infty} c_n g_n(z), \quad |z| < s,$$

there is a number  $A(s)$ , not depending on  $f(z)$ , such that

$$(3.12) \quad \sum_{n=0}^{\infty} |c_n|^2 s^{2n} \leq A(s) \int_0^{2\pi} |f(se^{i\theta})|^2 d\theta.$$

This follows from the last part of Theorem 2.1.

**4. Improvement of the criteria.** The conditions established in §3 can be generalized by restricting the  $h_n(z)$  only for large  $n$ . The generalized conditions occur in part in the literature, and in part are new. It turns out that only the behavior of the  $h_n(z)$  for large  $n$  is relevant to the existence of expansions of the type which we consider, as the following lemma shows.

**LEMMA.** Suppose that  $g_n(z)$  and  $g_n^*(z)$  are analytic in  $|z| < r_0$ , and  $g_n(z) \equiv g_n^*(z)$  for  $n > N$ . If  $\{g_n(z)\}$  has Property T in  $|z| < s_1 < r_0$ , and  $\{g_n^*(z)\}$  has Property T in every circle  $|z| < s^* \leq s_2$ , where  $r_0 \geq s_2 > s_1$ , then  $\{g_n(z)\}$  has Property T in  $|z| < s_2$ .

Let  $F(z)$  be an arbitrary function analytic in  $|z| < s_2$  and continuous in  $|z| \leq s_2$ , and let

$$(4.1) \quad F(z) = \sum_{n=0}^{\infty} c_n g_n(z),$$

where the series converges uniformly in any circle  $|z| \leq s'_1 < s_1$ . Define a function  $G(z)$  by the relation

$$(4.2) \quad G(z) = F(z) - \sum_{n=0}^N c_n g_n(z) = \sum_{n=N+1}^{\infty} c_n g_n(z) = \sum_{n=N+1}^{\infty} c_n g_n^*(z).$$

Now  $G(z)$  has a unique expansion of the form

$$(4.3) \quad G(z) = \sum_{n=0}^{\infty} d_n g_n^*(z),$$

the series converging uniformly in  $|z| \leq s'_2 < s_2$ . By comparison with (4.2), we see that  $d_n = 0$  ( $n = 0, 1, 2, \dots, N$ ). Hence the series in (4.1) converges uniformly in  $|z| \leq s'_2$ , and necessarily converges to  $F(z)$ . Since  $s'_2$  is any number less than  $s_2$ , the proof is complete.

We now suppose, as in §3, that  $g_n(z) = z^n [1 + h_n(z)]$ , where

$$h_n(z) = \sum_{k=1}^{\infty} \gamma_k^{(n)} z^k \quad (n = 0, 1, 2, \dots).$$

We suppose further that

$$|\gamma_k^{(n)}| \leq \delta_k^{(n)} \quad (k = 1, 2, \dots; n = 0, 1, 2, \dots),$$

where the series

$$H_n(z) = \sum_{k=1}^{\infty} \delta_k^{(n)} z^k \quad (n = 0, 1, 2, \dots)$$

converge in  $|z| < r_0$ . We introduce, for  $\rho < r_0$ , the quantities

$$K_{\rho,n}^2 = \sum_{k=1}^{\infty} (\delta_k^{(n)} \rho^k)^2, \quad L_{\rho}^{(n)} = \sup_{1 \leq k < \infty} \delta_k^{(n)} \rho^k,$$

and we then set

$$(4.4) \quad K_{\rho} = \limsup_{n \rightarrow \infty} K_{\rho,n}, \quad L_{\rho} = \limsup_{n \rightarrow \infty} L_{\rho}^{(n)}, \quad h(r) = \limsup_{n \rightarrow \infty} H_n(r) \quad (0 < r < r_0).$$

Then we can state

**THEOREM 4.1.** *The functions  $g_n(z)$  have Property T in any circle  $|z| < s$  provided that  $s$  satisfies one of the following three conditions<sup>(12)</sup>:*

$$(4.5) \quad h(s) < 1,$$

$$(4.6) \quad s < \sup_{0 \leq \rho < r_0} \frac{\rho}{(K_{\rho}^2 + 1)^{1/2}},$$

$$(4.7) \quad s < \sup_{0 \leq \rho < r_0} \frac{\rho}{L_{\rho} + 1}.$$

Theorem 4.1 states that Theorems 3.1 and 3.3 remain valid when  $K_{\rho}$ ,  $L_{\rho}$ , and  $h(r)$  are defined by (4.4). The theorem follows at once from the lemma, with  $g_n^*(z) = z^n$  for  $n < N$ , where  $N$  is chosen sufficiently large. It is only necessary to verify that the functions  $g_n(z)$  have Property T in some circle  $|z| < s_1$ . An application of Theorem 3.1 shows at once that this is true, with (for example)  $s_1$  such that

$$\sup_{0 \leq n < \infty} H_n(s_1) < 1.$$

Alternatively, we may suppose that, for  $n = 0, 1, 2, \dots$ ,

$$|h_n(z)| \leq M_n(\rho) \quad (|z| \leq \rho),$$

and that

$$(4.8) \quad M(\rho) = \limsup_{n \rightarrow \infty} M_n(\rho)$$

is finite. Then we can state

<sup>(12)</sup> For the theorem under (4.7), see G. S. Ketchum [4].

THEOREM 4.2<sup>(12)</sup>. The conclusion of Theorem 3.2 holds if the quantity  $M(\rho)$  in (3.9) is defined by (4.8).

Expansions in terms of functions  $g_n(z)$ , analytic in  $|z| < r_0$ , are particularly interesting if the expansion of every  $f(z)$  which is analytic in  $|z| < s$  converges in every circle  $|z| \leq s' < s$  (where naturally  $s \leq r_0$ ). This property (which we may call Property U) is possessed, of course, by the functions  $z^n$ . Using the theorems of this section, we can easily obtain the following sufficient conditions for a set  $g_n(z)$  to have Property U:

$$(4.9) \quad h(r_0) < 1,$$

$$(4.10) \quad L_{r_0} = 0,$$

$$(4.11) \quad M(r_0) = 0,$$

where  $h(r_0)$ ,  $L_{r_0}$ , and  $M(r_0)$  denote the limits of the respective functions of  $r$  (defined in (4.4) and (4.8)) as  $r \rightarrow r_0$ . Condition (4.11) shows, for example, that if

$$(4.12) \quad h_n(z) = o(1), \quad n \rightarrow \infty,$$

uniformly with respect to  $z$  in each circle  $|z| \leq r' < r_0$ , then the set  $\{g_n(z)\} = \{z^n [1 + h_n(z)]\}$  has Property U. This result was obtained by Sheffer and by Takahashi<sup>(14)</sup>; it generalizes a result of Widder [20], in which the condition  $h_n(z) = O(1/n)$  appears instead of (4.12). Condition (4.9) will sometimes establish Property U when (4.12) is not satisfied. For example, if

$$h_n(z) = \sum_{k=1}^{\infty} \delta_k^{(n)} z^k,$$

and

$$|\delta_k^{(n)}| \leq \frac{1}{k(k+1)} \quad (n = 0, 1, \dots; k = 1, 2, \dots),$$

we have

$$h(z) = 1 - \frac{1-z}{z} \log(1-z) = \sum_{k=1}^{\infty} \frac{z^k}{k(k+1)},$$

and  $h(r) < 1$  if  $r < 1$ . In this case the corresponding functions  $g_n(z)$  have Property U in  $|z| < 1$ , although neither (4.10), (4.11), nor (4.12) is necessarily satisfied.

**5. Further generalizations.** In Theorems 4.1 and 4.2 we made restrictions on the individual coefficients in the power series of the functions  $h_n(z)$ . In this section a method will be developed for replacing such restrictions by re-

<sup>(12)</sup> Takenaka [15], G. S. Ketchum [4].

<sup>(14)</sup> Sheffer [10, pp. 588, 597], Takahashi [13]. Cf. G. S. Ketchum [4, p. 215].

strictions on linear combinations of the coefficients. The results obtained (which could evidently be generalized still further) overlap those of G. S. Ketchum [4], who first obtained such results.

Suppose that  $L$  is a one-to-one linear operation from  $H_2(r)$  to  $H_2(r)$  (so that  $L$  has a linear inverse). If the set of functions  $G_n(z) = L[g_n(z)]$  is a base, and  $f(z)$  is an arbitrary element of  $H_2(r)$ , we have a unique expansion

$$L[f(z)] = \sum_{n=0}^{\infty} a_n G_n(z);$$

since  $L^{-1}$  is continuous, we have

$$f(z) = \sum_{n=0}^{\infty} a_n g_n(z);$$

if we also have

$$f(z) = \sum_{n=0}^{\infty} b_n g_n(z),$$

then

$$L[f(z)] = \sum_{n=0}^{\infty} b_n G_n(z),$$

and  $a_n = b_n$  ( $n = 0, 1, 2, \dots$ ). The convergence is convergence in the topology of  $H_2(r)$ ; this, as we have seen, implies uniform convergence in every circle  $|z| \leq r' < r$ . Hence an expansion theorem for the functions  $L[g_n(z)]$  gives rise to an expansion theorem for the  $g_n(z)$  themselves. A trivial, but not unimportant, illustration is given by the operator  $L$  which transforms  $g_n(z)$  into  $\sigma(z)g_n(z)$ , where  $\sigma(z)$  is analytic and bounded in  $|z| \leq r$ ,  $\sigma(0) = 1$ , and  $\sigma(z) \neq 0$  in  $|z| \leq r$ .

We now discuss a case which is not entirely covered by the procedure just outlined; it includes some of the results of G. S. Ketchum mentioned above. For simplicity we consider only the special case when the coefficients of the functions  $h_n(z)$  are combined two at a time. Let  $\{k_r\}$  ( $r = 1, 2, \dots$ ) be a sequence of complex numbers such that

$$\limsup_{r \rightarrow \infty} |k_r|^{1/r} \leq 1,$$

so that  $\Lambda(z) = \sum_{r=0}^{\infty} k_r z^r$  is analytic in  $|z| < 1$ . If

$$f(z) = \sum_{r=0}^{\infty} b_r z^r, \quad |z| < r,$$

we define

$$(5.1) \quad F(z) = L[f] = f(z) + \sum_{r=0}^{\infty} k_r b_{r+1} z^r = \sum_{r=0}^{\infty} (b_r + k_r b_{r+1}) z^r. \quad (b_{-1} = 0)$$



By Hadamard's multiplication theorem<sup>(15)</sup>,  $F(z)$  has no singularities inside the circle  $|z| < r$ . We have the representation

$$(5.2) \quad F(z) = f(z) - \frac{1}{2\pi i} \int_C \Lambda(z/w) f(w) dw \quad (|z| < r' < r)$$

where  $C$  is the circle  $|w| = r' < r$ .

Let us now consider the expansion of a given analytic function  $f(z)$  in terms of a set of functions

$$g_n(z) = z^n [1 + h_n(z)],$$

with  $h_n(z)$  analytic in  $|z| < r$  and  $h_n(0) = 0$ . We introduce the functions

$$(5.3) \quad F(z) = L[f], \quad G_n(z) = L[g_n];$$

it is clear that we have  $G_n(z) = z^n [1 + H_n(z)]$ , where the  $H_n(z)$  are analytic in  $|z| < r$  and  $H_n(0) = 0$ . (It is not in general true that  $G_n(z) \in H_2(r)$  if  $g_n(z) \in H_2(r)$ ). Let us suppose that the  $G_n(z)$  satisfy one of the conditions of Theorems 4.1 and 4.2, so that we have in  $|z| < r$  a unique expansion

$$F(z) = \sum_{n=0}^{\infty} c_n G_n(z)$$

converging uniformly in any circle  $|z| \leq s < r$ , with

$$\sum_{n=0}^{\infty} |c_n|^2 s^{2n} \leq A(s) \int_0^{2\pi} |F(se^{i\theta})|^2 d\theta;$$

the last relation follows from the remark made at the end of §3. From (5.2) and (5.3) we then have

$$(5.4) \quad f(z) - \frac{1}{2\pi i} \int_C \Lambda(z/w) f(w) dw = \sum_{n=0}^{\infty} c_n \left\{ g_n(z) - \frac{1}{2\pi i} \int_C \Lambda(z/w) g_n(w) dw \right\},$$

the series converging in  $|z| < r$ , uniformly in  $|z| \leq s < r$ . If  $s$  is temporarily fixed, and we take  $\rho$  so that  $s < \rho < r$ , the series  $\sum |c_n|^2 \rho^{2n}$  is convergent, and the functions  $w^{-n} g_n(w)$  are uniformly bounded on  $|w| = s$ . It follows, by an application of Cauchy's inequality, that the series  $\sum c_n g_n(w)$  is uniformly (and absolutely) convergent on  $|w| = s$  and so in  $|w| \leq s$ . Since the series on the right of (5.4) is uniformly convergent in any circle  $|z| \leq s < r$ , we have

$$f(z) - \sum_{n=0}^{\infty} c_n g_n(z) - \frac{1}{2\pi i} \int_C \Lambda(z/w) \left\{ f(w) - \sum_{n=0}^{\infty} c_n g_n(w) \right\} dw \equiv 0$$

in  $|z| < r$ , both series being uniformly convergent in any circle  $|z| \leq s < r$ . That is, the function  $f^*(z)$  defined by

<sup>(15)</sup> See, e.g., Dienes [24, p. 346].

$$f^*(z) = f(z) - \sum_{n=0}^{\infty} c_n g_n(z) = \sum_{\nu=0}^{\infty} b_{\nu}^* z^{\nu} \quad (b_{-1}^* = 0)$$

is analytic in  $|z| < r$ , and we have  $L[f^*(z)] \equiv 0$  in  $|z| < r$ . From (5.1) we see that this means that

$$b_{\nu}^* + k_{\nu} b_{\nu-1}^* = 0 \quad (\nu = 0, 1, 2, \dots),$$

and hence that  $b_{\nu}^* = 0$  ( $\nu = 0, 1, 2, \dots$ ). That is,

$$f(z) = \sum_{n=0}^{\infty} c_n g_n(z),$$

the series converging uniformly in any circle  $|z| \leq s < r$ . We sum up our conclusions in a formal theorem.

**THEOREM 5.1.** *If the functions  $G_n(z) = L[g_n(z)]$ , where  $L$  is defined by (5.1), satisfy the conditions of Theorem 4.1 or Theorem 4.2, and  $s$  is defined as in those theorems, the functions  $g_n(z)$  have Property T in  $|z| < s$ .*

For example, if the numbers  $k_{\nu}$  satisfy

$$\limsup_{\nu \rightarrow \infty} |k_{\nu}|^{1/\nu} \leq 1;$$

if

$$h_n(z) = \sum_{\nu=1}^{\infty} \gamma_{\nu}^{(n)} z^{\nu} \quad (|z| < r_0),$$

$$|\gamma_{\nu}^{(n)} + k_{n+\nu} \gamma_{\nu-1}^{(n)}| \leq \beta_{\nu} \quad (n = 0, 1, \dots; \nu = 1, 2, \dots),$$

$$h(z) = \sum_{\nu=1}^{\infty} \beta_{\nu} z^{\nu} \quad (|z| < r_0),$$

and  $h(s) < 1$ , then the functions  $z^n [1 + h_n(z)]$  have Property T in  $|z| < s$ .

It is clear that linear operations other than that defined in (5.1) could also be used.

## 6. Special expansion theorems.

**THEOREM 6.1.** *If  $\phi(z)$  is an analytic function whose Maclaurin series has positive coefficients<sup>(16)</sup>, and radius of convergence  $R$  ( $R \leq \infty$ ), if  $\phi(0) = 1$ , and if the complex numbers  $\alpha_n$  satisfy*

$$(6.1) \quad \limsup_{n \rightarrow \infty} |\alpha_n| \leq 1, \quad \sup_{0 \leq n < \infty} |\alpha_n| < R/\phi^{-1}(2),$$

the functions

$$(6.2) \quad z^n \phi(\alpha_n z) \quad (n = 0, 1, 2, \dots)$$

have Property T in any circle  $|z| \leq s < \phi^{-1}(2)$ .

<sup>(16)</sup> That is,  $\phi(z)$  is absolutely monotonic on the segment  $(0, R)$  of the real axis.

Here we have  $h_n(z) = \phi(\alpha_n z) - 1$ ; and, in the notation of §4,

$$H_n(r) = \phi(|\alpha_n| r) - 1, \quad h(r) \leq \phi(r) - 1,$$

and  $h(r) < 1$  if  $\phi(r) < 2$ .

In particular, we may have<sup>(17)</sup>  $\phi(z) = e^z$ .

Various modifications of the situation considered in Theorem 6.1 are possible. We shall discuss three which have interesting applications.

**THEOREM 6.2.** *The functions  $g_n(z)$  defined by*

$$g_0(z) = 1,$$

$$g_n(z) = z^{n-1} \frac{e^{\alpha_n z} - e^{\beta_n z}}{\alpha_n - \beta_n}$$

$$(|\alpha_n| \leq 1, |\beta_n| \leq 1, \alpha_n \neq \beta_n; n = 1, 2, \dots)$$

have Property T in the circle  $|z| < r$  if  $r < \log 2$ .

It is easy to show that we may take  $h(r) = e^r - 1$  in this case. For details, the reader is referred to the author's note [22] where Theorem 6.2 is applied to show that an entire function of exponential type less than  $\log 2$  has an infinite number of derivatives which are univalent in the unit circle, unless it is a polynomial.

For the next two theorems, it is necessary to go back to Theorem 2.2.

**THEOREM 6.3.** *The functions  $g_n(z)$  defined by*

$$g_{2n}(z) = z^{2n} e^{\alpha_n z} \quad (|\alpha| \leq 1),$$

$$g_{2n+1}(z) = z^{2n+1},$$

have Property T in any circle  $|z| < r < 0.780$ .

We note that  $\log 2 = 0.693$ ,  $\pi/4 = 0.785$ . We thus have more than Theorem 6.1 would establish, but still less than the result which may be conjectured<sup>(18)</sup>, that Theorem 6.1 holds, when  $\phi(z) = e^z$ , for  $s < \pi/4$ .

By Theorem 2.2, Theorem 6.3 will follow if we show that for every sequence  $\{a_n\}$  of complex numbers and for every  $N$

$$(6.3) \quad \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^N a_n [g_n(re^{i\theta}) - r^n e^{in\theta}] \right|^2 d\theta \leq \lambda(r) \sum_{n=0}^N |a_n|^2 r^{2n}, \quad r < 0.780,$$

with  $\lambda(r) < 1$ . The left-hand side, by the reasoning of Theorem 3.1, does not exceed

<sup>(17)</sup> The corresponding theorem, with region of convergence  $|z| < 1/e$ , was given by Takemaka [15]. See also footnote 2.

<sup>(18)</sup> See footnote 21.

$$\begin{aligned}
& \frac{1}{2\pi} \int_0^{2\pi} |e^z - 1|^2 \left| \sum_{n=0}^{[N/2]} a_{2n} z^{2n} \right|^2 d\theta \quad (z = re^{i\theta}) \\
&= \frac{1}{2\pi} \int_0^{2\pi} |e^z - 1|^2 |\psi(z)|^2 d\theta \\
&= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} |\psi(z)(e^z - 1)|^2 d\theta + \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} |\psi(z)(e^{-z} - 1)|^2 d\theta,
\end{aligned}$$

since  $|\psi(z)|$  is periodic in  $\theta$  with period  $\pi$ . Thus the left side of (6.3) does not exceed

$$\begin{aligned}
& \frac{1}{2\pi} \left\{ (e^r - 1)^2 + \max_{-\pi/2 \leq \theta \leq \pi/2} |e^{-z} - 1|^2 \right\} \int_{-\pi/2}^{\pi/2} |\psi(z)|^2 d\theta \\
&= \frac{1}{2} \left\{ (e^r - 1)^2 + \max_{-\pi/2 \leq \theta \leq \pi/2} |e^{-z} - 1|^2 \right\} \sum_{n=0}^{[N/2]} |a_{2n}|^2 r^{4n}.
\end{aligned}$$

We have

$$\begin{aligned}
|e^z - 1|^2 &= e^{-2r \cos \theta} - 2e^{-r \cos \theta} \cos(r \sin \theta) + 1 \\
&\leq e^{-2r \cos \theta} + 1 - 2e^{-r \cos \theta} \cos r \\
&= A(\theta),
\end{aligned}$$

say. Now

$$A'(\theta) = 2r \sin \theta e^{-r \cos \theta} (e^{-r \cos \theta} - \cos r),$$

and vanishes only when  $\theta=0$  or when  $\cos \theta = (1/r) \log \cos(1/r)$ . In the latter case,  $\exp(-r \cos \theta) = \cos r$ , and

$$A(\theta) = 1 - \cos^2 r < 2(1 - \cos r) = A(\frac{1}{2}\pi).$$

Consequently  $A(\theta)$  assumes its maximum when  $\theta=0$  or  $\theta=\frac{1}{2}\pi$ . For  $r=0.780$ , we find that  $A(0) < A(\frac{1}{2}\pi)$ . In fact, this inequality is

$$e^{-2r} - 2e^{-r} + 1 < 2(1 - \cos r) + 2e^{-r} \cos r - 2e^{-r},$$

which is equivalent to

$$1 + e^{-r} > 2 \cos r,$$

which is satisfied for  $r=0.780$ . Hence

$$A(\theta) \leq A(\frac{1}{2}\pi) = 2(1 - \cos r),$$

and the left side of (6.3) does not exceed

$$\frac{1}{2} \{ (e^r - 1)^2 + 2(1 - \cos r) \} \sum_{n=0}^N |a_n|^2 r^{2n};$$

the brace is less than 2 when  $r=0.780$ . This completes the proof of Theorem 6.3.

THEOREM 6.4. If  $r < \log 2$ , and the complex numbers  $\beta_n$  are such that

$$(6.4) \quad \sum_{n=0}^{\infty} |\beta_n| 2r^{2n} < 2e^r - e^{2r},$$

the functions  $g_n(z)$  defined by

$$(6.5) \quad g_n(z) = z^n e^{a_n z} - \beta_n \quad (|\alpha_n| \leq 1)$$

have Property T in  $|z| < r$ .

To apply Theorem 2.2, we need to show that if  $h_n(z) = e^{a_n z} - 1$ , then for all  $\{a_n\}$  and  $N$ ,

$$(6.6) \quad \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^N a_n [z^n h_n(z) - \beta_n] \right|^2 d\theta \leq \lambda(r) \sum_{n=0}^N |a_n|^2 r^{2n} \quad (z = re^{i\theta}),$$

with  $\lambda(r) < 1$ , when  $r$  and  $\{\beta_n\}$  satisfy (6.4).

The left side of (6.6) may be written in the form

$$\begin{aligned} & \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^N a_n z^n h_n(z) - \sum_{n=0}^N a_n \beta_n \right|^2 d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^N a_n z^n h_n(z) \right|^2 d\theta - \frac{1}{\pi} \int_0^{2\pi} \Re \left\{ \sum_{n=0}^N a_n z^n h_n(z) \sum_{m=0}^N \bar{a}_m \bar{\beta}_m \right\} d\theta \\ & \quad + \left| \sum_{n=0}^N a_n \beta_n \right|^2 \\ &= S_1 - S_2 + S_3. \end{aligned}$$

Now, by the proof of Theorem 3.1,

$$(6.7) \quad S_1 \leq (e^r - 1)^2 \sum_{n=0}^N |a_n|^2 r^{2n};$$

and

$$(6.8) \quad S_3 \leq \left( \sum_{n=0}^N |a_n| r^n |\beta_n| r^{-n} \right)^2 \leq \sum_{n=0}^N |a_n|^2 r^{2n} \sum_{n=0}^N |\beta_n|^2 r^{-2n}$$

by Cauchy's inequality. Finally,

$$S_2 = 2\Re \left\{ \frac{1}{2\pi} \int_0^{2\pi} \sum_{n=0}^N a_n z^n h_n(z) \sum_{m=0}^N \bar{a}_m \bar{\beta}_m \frac{dz}{iz} \right\} = 0,$$

since  $h_n(0) = 0$ . Combining this with (6.7) and (6.8), we have

$$\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{n=0}^N a_n [z^n h_n(z) - \beta_n] \right|^2 d\theta \leq \left\{ (e^r - 1)^2 + \sum_{n=0}^N |\beta_n|^2 r^{-2n} \right\} \sum_{n=0}^N |a_n|^2 r^{2n}.$$



Thus by Theorem 2.2 the system (6.5) has Property T if the brace in the last inequality is less than one. This will clearly be true if (6.4) is satisfied.

**7. Applications.** We now use the theorems of §6 to prove theorems concerning the values taken by derivatives of entire functions of order one and exponential type. We need the following lemma.

**LEMMA<sup>(19)</sup>.** *If  $f(z)$  is an entire function of exponential type  $k$ , it has the representation*

$$(7.1) \quad f(z) = \int_C e^{zw} F(w) dw,$$

where  $C$  is any circle  $|z| = k' > k$ , and  $F(w)$  is analytic outside  $|z| = k$ .

It follows that

$$(7.2) \quad f^{(n)}(z) = \int_C w^n e^{zw} F(w) dw \quad (n = 1, 2, \dots).$$

If now the functions  $g_n(w)$  have property T in  $|w| \leq k'$ , we can expand the function  $e^{zw}$  in terms of them, substitute in (7.1), and integrate term by term<sup>(20)</sup>. We thus obtain an expansion of the form

$$(7.3) \quad f(z) = \sum_{n=0}^{\infty} c_n(z) \int_C g_n(w) F(w) dw.$$

We can now establish the following theorems.

**THEOREM 7.1.** *If  $f(z)$  is an entire function of exponential type  $k < \log 2$ , and if  $f(0) = 1$ ,  $|\alpha_n| \leq 1$ , and  $r < k$ , the inequality*

$$(7.4) \quad \sum_{n=0}^{\infty} |f^{(n)}(\alpha_n)|^2 r^{-2n} \geq 2e^r - e^{2r}$$

is valid.

If (7.4) is not true, Theorem 6.4 applies, with  $\beta_n = f^{(n)}(\alpha_n)$ , and (7.3) has the form

$$f(z) = \sum_{n=0}^{\infty} c_n(z) \{f^{(n)}(\alpha_n) - \beta_n f(0)\} = 0,$$

which is impossible since  $f(0) = 1$ .

As a corollary we obtain the following theorem of S. Takenaka [29].

**THEOREM 7.2.** *If  $f(z)$  is an entire function of exponential type  $k < \log 2$ , and<sup>(21)</sup>  $|\alpha_n| \leq 1$ , then*

<sup>(19)</sup> See Pólya [27, pp. 580 ff.].

<sup>(20)</sup> Cf. Whittaker [30, p. 67], Gelfond [25].

<sup>(21)</sup> Or even if  $\limsup |\alpha_n| \leq 1$ .

$$(7.5) \quad f^{(n)}(\alpha_n) = 0 \quad (n = 0, 1, 2, \dots)$$

implies  $f(z) \equiv 0$ .

For, if  $f(0) \neq 0$ , we consider  $f(z)/f(0)$ , to which Theorem 7.1 applies, since the left side of (7.4) is zero if (7.5) is satisfied. If  $f(0) = 0$ , while  $f(z_0) \neq 0$  for some  $z_0$  in  $|z| < 1$ , we apply Theorem 7.1 to the function  $[f(z_0) - f(z)]/f(z_0)$ , taking  $\alpha_0 = z_0$ .

From Theorem 6.3, we obtain

**THEOREM 7.3.** *If  $f(z)$  is an entire function of exponential type  $k < 0.780$ , and  $|\alpha_n| \leq 1$ , then the conditions*

$$f^{(2n+1)}(0) = f^{(2n)}(\alpha_n) = 0 \quad (n = 0, 1, 2, \dots)$$

imply that  $f(z) \equiv 0$ .

This is more than follows from Theorem 7.2, but less than would follow if Theorem 7.2 were proved to be true with  $k < \pi/4$ , as has been conjectured<sup>(22)</sup>.

Analogous theorems concerning functions analytic in a finite circle<sup>(23)</sup> can be proved by developing  $(w-z)^{-1}$ , as a function of  $w$ , in terms of (for example)

$$\frac{w^n}{(1 - \alpha_n w)^n},$$

and substituting the expansion into Cauchy's integral formula.

#### PAPERS ON GENERAL EXPANSION THEOREMS

1. G. D. Birkhoff, *Sur une généralisation de la série de Taylor*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 163 (1917), pp. 942-945.
2. R. F. Graesser, *A certain general type of Neumann expansions and expansions in confluent hypergeometric functions*, American Journal of Mathematics, vol. 49 (1927), pp. 577-597.
3. S. Izumi, *On the expansion of analytic function*, Tôhoku Mathematical Journal, vol. 28 (1927), pp. 97-106.
4. G. S. Ketchum, *On certain generalizations of the Cauchy-Taylor expansion theory*, these Transactions, vol. 40 (1936), pp. 208-224.
5. P. W. Ketchum, *Infinite systems of linear equations and expansions of analytic functions*, Duke Mathematical Journal, vol. 4 (1938), pp. 668-677.
6. T. Kubota, *Eine Verallgemeinerung des Taylor-Cauchyschen Satzes*, Tôhoku Mathematical Journal, vol. 22 (1923), pp. 336-347.
7. S. Narumi, *A theorem on the expansion of analytic functions*, Tôhoku Mathematical Journal, vol. 30 (1929), pp. 441-444.
8. Y. Okada, *On a certain expansion of analytic function*, Tôhoku Mathematical Journal, vol. 22 (1923), pp. 325-335.
9. S. Pincherle, *Sopra alcuni sviluppi in serie per funzioni analitiche*, Memorie della Reale Accademia delle Scienze dell'Istituto di Bologna, (4), vol. 3 (1881), pp. 151-180.

<sup>(22)</sup> See Whittaker [30, p. 45], Schoenberg [28].

<sup>(23)</sup> See Takenaka [15, 29]; Whittaker [30, p. 43].

10. I. M. Sheffer, *Concerning some methods of best approximation, and a theorem of Birkhoff*, American Journal of Mathematics, vol. 57 (1935), pp. 587-614.
11. S. Takahashi, *On the expansion of analytic function*, Proceedings of the Imperial Academy, Tokyo, vol. 6 (1930), pp. 389-392.
12. ———, *A remark on Mr. D. V. Widder's theorem*, Tôhoku Mathematical Journal, vol. 33 (1930), pp. 48-54.
13. ———, *On the expansion of analytic functions*, Tôhoku Mathematical Journal, vol. 35 (1932), pp. 242-243.
14. S. Takenaka, *A generalization of Taylor's series*, Japanese Journal of Mathematics, vol. 7 (1930), pp. 187-198.
15. ———, *On the expansion of analytic functions in series of analytic functions and its application to the study of the distribution of zero points of the derivatives of analytic functions*, Nippon Sôgaku-Buturigakkwai Kizi (Proceedings of the Physico-Mathematical Society of Japan), (3), vol. 13 (1931), pp. 111-132.
16. L. Tonelli, *Sulle serie di funzioni analitiche della forma  $\sum a_n(x)x^n$* , Annali di Matematica Pura ed Applicata, (3), vol. 18 (1911), pp. 99-103.
17. J. L. Walsh, *On the expansion of analytic functions in series of polynomials*, these Transactions, vol. 26 (1924), pp. 155-170.
18. ———, *On the expansion of analytic functions in series of polynomials and in series of other analytic functions*, these Transactions, vol. 30 (1928), pp. 307-332.
19. ———, *Note on the expansion of analytic functions in series of polynomials and in series of other analytic functions*, these Transactions, vol. 31 (1929), pp. 53-57.
20. D. V. Widder, *On the expansion of analytic functions of a complex variable in generalized Taylor's series*, these Transactions, vol. 31 (1929), pp. 43-52.

## OTHER REFERENCES

21. S. Banach, *Théorie des Opérations Linéaires*, 1932.
22. R. P. Boas, Jr., *Univalent derivatives of entire functions*, Duke Mathematical Journal, vol. 6 (1940), pp. 719-721.
23. H. F. Bohnenblust and A. Sobczyk, *Extensions of functionals on complex linear spaces*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 91-93.
24. P. Dienes, *The Taylor Series, An Introduction to the Theory of Functions of a Complex Variable*, 1931.
25. A. Gelfond, *Interpolation et unicité des fonctions entières*, Matematiceskii Sbornik (Recueil Mathématique), new series, vol. 4 (1938), pp. 115-147.
26. R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 19, 1934.
27. G. Pólya, *Untersuchungen über Lücken und Singularitäten von Potenzreihen*, Mathematische Zeitschrift, vol. 29 (1929), pp. 549-640.
28. I. J. Schoenberg, *On the zeros of successive derivatives of integral functions*, these Transactions, vol. 40 (1936), pp. 12-23.
29. S. Takenaka, *On the expansion of integral transcendental functions in generalized Taylor's series*, Nippon Sôgaku-Buturigakkwai Kizi (Proceedings of the Physico-Mathematical Society of Japan), (3), vol. 14 (1932), pp. 529-542.
30. J. M. Whittaker, *Interpolatory Function Theory*, 1935.
31. A. Zygmund, *Trigonometrical Series*, 1935.
32. I. I. Ibragimoff (I. Ibragimoff), *Sur quelques systèmes complets de fonctions analytiques* (in Russian), Izvestiya Akademii Nauk SSSR, Seriya Matematicheskaya (Bulletin de l'Académie des Sciences de l'URSS, Série Mathématique), 1939, pp. 553-567; French summary, pp. 567-568.

DUKE UNIVERSITY,  
DURHAM, N. C.

# ON THE INTEGRO-DIFFERENTIAL EQUATIONS OF PURELY DISCONTINUOUS MARKOFF PROCESSES

BY  
WILLY FELLER

1. **Introduction.** In the following we are concerned with stochastic processes depending on a continuous time parameter  $t$ , that is to say, with some entity (chance variable) whose state is specified by a point  $X(t)$  varying in some space  $E$  according to some probability law. The process is called a *Markoff process*<sup>(1)</sup> if the probability distribution of  $X(t)$  is completely determined for all  $t > \tau$  by the knowledge of the state  $X(\tau)$ ; and in particular is independent of the development of the process for  $t < \tau$ <sup>(2)</sup>. Analytically a Markoff process is completely determined by its transition probabilities  $P(\tau, x; t, \Delta)$ , giving the conditional probability of  $X(t)$ 's being contained, at the moment  $t$ , in the set  $\Delta \subset E$  under the hypothesis that at a fixed moment  $\tau < t$  the state  $X(\tau)$  coincided with the point  $x$  of  $E$ .

In strict terms, we shall suppose that there is specified, in the space  $E$ , a Borel field  $\mathfrak{B}$  of sets (on which probabilities are defined) such that  $E \in \mathfrak{B}$  and also any set consisting of a single point belongs to  $\mathfrak{B}$ . It is then required that  $P(\tau, x; t, \Delta)$  is, for fixed  $\tau, t > \tau$  and  $x \in E$ , a non-negative and completely additive function of sets on  $\mathfrak{B}$ , with

$$(1) \quad P(\tau, x; t, E) = 1.$$

Moreover, we shall always assume that for fixed  $\tau, t, \Delta$  the function  $P(\tau, x; t, \Delta)$  is measurable with respect to  $\mathfrak{B}$ , that is to say, that for any  $a > 0$  the set where  $P(\tau, x; t, \Delta) < a$  belongs to  $\mathfrak{B}$ . Finally we shall, for the sake of simplicity, restrict ourselves to  $P(\tau, x; t, \Delta)$  depending, for fixed other arguments, continuously on both  $\tau$  and  $t$ <sup>(3)</sup>. This implies in particular that as either  $t \rightarrow \tau + 0$  or  $\tau \rightarrow t - 0$

---

Presented to the Society, February 24, 1940; received by the editors March 5, 1940.

(<sup>1</sup>) This name was chosen in accordance with the now common terminology in the case of processes with an integral-valued parameter  $t$ . Kolmogoroff [6] calls such processes stochastically definite, and this terminology I had also adopted previously. Markoff processes are, sometimes, also described as being "without after effect," or as being submitted to an "influence globale" (Pólya).

(<sup>2</sup>) This is, of course, not meant to be a strict definition; as a matter of fact, we shall be concerned only with the function  $P(\tau, x; t, \Delta)$ , which will be defined purely analytically.

(<sup>3</sup>) It may be pointed out that this does not imply the continuity of the movement of  $X(t)$ . We shall, on the contrary, be concerned only with states  $X(t)$  changing abruptly by jumps. The continuity of  $P(\tau, x; t, \Delta)$  means that the probability of a jump during a small time-interval is small.

$$(2) \quad P(\tau, x; t, \Lambda) \rightarrow \delta(x, \Lambda) = \begin{cases} 1 & \text{if } x \in \Lambda, \\ 0 & \text{otherwise.} \end{cases}$$

Subdividing now the interval  $(\tau, t)$  by a point  $s$  and considering all possible states  $X(s)$ , we readily get the identity

$$(3) \quad P(\tau, x; t, \Lambda) = \int_E P(\tau, x; s, dE_y) P(s, y; t, \Lambda)$$

known as the equation of Chapman and Kolmogoroff<sup>(4)</sup>. We shall take these relations as the analytic definition of a Markoff process and consider any  $P(\tau, x; t, \Lambda)$  of the described sort as defining the transition probabilities of such a process<sup>(5)</sup>.

In the special case of the space  $E$  containing at most an enumerable number of points, we shall denote these points by  $x_k$  and write

$$(4) \quad P(\tau, x_i; t, x_k) = P_{ik}(\tau, t).$$

By (1) we have  $\sum_k P_{ik}(\tau, t) = 1$ , while (2) and (3) are respectively equivalent to

$$(5) \quad P_{ik}(\tau, t) \rightarrow \delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} \quad (t - \tau \rightarrow +0),$$

and

$$(6) \quad P_{ik}(\tau, t) = \sum_j P_{ij}(\tau, s) P_{jk}(s, t) \quad (\tau < s < t).$$

Now a *purely discontinuous process* may be described by the following property: if, at the moment  $t$ , the actual state is given by the point  $x$ , then there is a probability  $1 - p(t, x)\Delta t + o(\Delta t)$  that no change of state will occur during  $(t, t + \Delta t > t)$ ; and if a change occurs, the probability of  $X(t + \Delta t)$ 's being contained in the set  $\Lambda$  is given, except for terms of  $o(1)$ , by a probability distribution  $\Pi(t, x, \Lambda)$ <sup>(6)</sup>. In strict terms we shall say that the *Markoff process*

<sup>(4)</sup> Cf. Kolmogoroff [6] where the foundations of the general theory of Markoff processes have been laid.

<sup>(5)</sup> This is the natural point of view for the purposes of the present paper. From an axiomatic point of view, however, any stochastic process corresponds to a measure in the space of real functions defined on  $E$ . Even in the case of Markoff processes there are problems (especially the problem of ruin, playing an important rôle in the theory of risk) which require a deeper penetration in the theory of the functional space. For the treatment of stochastic processes in terms of measure, the reader is referred to J. C. Doob's fundamental paper [1].

<sup>(6)</sup> Essentially this definition was given by Feller [3]; cf. also Dubrovski [2]. This kind of processes was mentioned also by Kolmogoroff [6].

Examples of such processes are furnished by the theory of radioactive processes and the theory of automatic telephone offices; by the transport of stones by rivers (treated by quite different methods of Pólya [8]); by the mathematical theory of struggle for life (Feller [4]), etc. Perhaps the most important application is furnished by the theory of risk.

There is no general definition for "purely continuous" processes in abstract  $E$ . In the Euclidean space such processes were defined by Kolmogoroff [6] and somewhat more generally by

defined by  $P(\tau, x; t, \Lambda)$  is purely discontinuous if for small  $t - \tau > 0$

$$(7) \quad P(\tau, x; t, \Lambda) = \{1 - p(t, x)(t - \tau)\} \delta(x, \Lambda) \\ + p(t, x)(t - \tau) \Pi(t, x, \Lambda) + o(t - \tau),$$

where  $\delta(x, \Lambda)$  was defined by (2) and the exact assumptions as to  $p(t, x)$  and  $\Pi(t, x, \Lambda)$  will be specified in §2, (i)–(ii); in general,  $o(t - \tau)$  will depend on  $x$  and  $\Lambda$ .

The main problem with which we are confronted is to determine whether or not to any two functions  $p(t, x)$  and  $\Pi(t, x, \Lambda)$ , subjected only to the conditions §2, (i)–(ii), there corresponds a Markoff process, whose transition probabilities  $P(\tau, x; t, \Lambda)$  satisfy (7); and if so, whether this process is uniquely determined.

It will be shown (§2) that there is a subclass  $\mathfrak{B}_1$ , of sets  $\Lambda \in \mathfrak{B}$  such that for all  $\Lambda \in \mathfrak{B}_1$ , all  $\tau$ , and almost all  $t$  the partial derivatives  $\partial P(\tau, x; t, \Lambda)/\partial t$  and  $\partial P(\tau, x; t, \Lambda)/\partial \tau$  exist; for all those  $\Lambda$ , all  $\tau$  and almost all  $t$ , the integro-differential equations

$$(8) \quad \frac{\partial P(\tau, x; t, \Lambda)}{\partial t} = - \int_{\Lambda} p(t, y) P(\tau, x; t, dE_y) \\ + \int_E p(t, y) \Pi(t, y, \Lambda) P(\tau, x; t, dE_y)$$

and

$$(9) \quad \frac{\partial P(\tau, x; t, \Lambda)}{\partial \tau} = p(\tau, x) \left\{ P(\tau, x; t, \Lambda) - \int_E P(\tau, y; t, \Lambda) \Pi(\tau, x, dE_y) \right\}$$

hold, implying the existence of the integrals for all  $\Lambda \in \mathfrak{B}_1$ , and almost all  $t$ . The class  $\mathfrak{B}_1$  contains, among others, sequences of sets  $\Lambda_n \uparrow E$ .

Thus the problem is reduced to the integration of (8)–(9). It will be shown (§§3–5) that there is a function  $P(\tau, x; t, \Lambda)$  which satisfies (8)–(9) for all  $\Lambda \in \mathfrak{B}_1$  and almost all  $t$  and  $\tau$ ; this  $P(\tau, x; t, \Lambda)$  is uniquely determined by each of the equations (8)–(9)<sup>(7)</sup> and has all properties described above, except perhaps (1); one has always

$$(10) \quad 0 \leq P(\tau, x; t, \Lambda) \leq 1,$$

but there are cases with

Feller [3]. This type is illustrated by the diffusion processes: there is a probability equal to 1 that some change of  $X(t)$  will occur during any time-interval, but the chance is near to 1 that the variation will be, in a specified sense, small for small intervals. This type is described by partial differential equations of parabolic type. There is also a "mixed type" leading to the equation (15) and its adjoint.

(<sup>7</sup>) It should be understood that, in general, a solution of (8) is not uniquely determined by the initial values (2), not even in the case of enumerable spaces. The uniqueness mentioned is a consequence of the additional hypothesis that  $0 \leq P(\tau, x; t, \Lambda) \leq 1$ .



$$(11) \quad P(\tau, x; t, E) < 1.$$

This exceptional case arises only if  $p(t, x)$  is unbounded (cf. Theorem 6), but can occur also in the case of enumerable spaces  $E$ .

The existence of positive bounded solutions that conform with all other requirements of the theory, including (3), but still fail to be distribution functions, is most striking, and an analysis of this phenomenon was the primary object, and constitutes the most delicate part, of the present investigation. In the case of temporally homogeneous processes, that is, in the case of  $p(t, x)$  and  $\Pi(t, x, \Delta)$  not depending on  $t$ , we give in §6 a *necessary and sufficient condition that the solution  $P(\tau, x; t, \Delta)$  be a proper probability distribution*, that is to say, that (1) holds. This condition is rather complicated, but can be interpreted in terms of the ergodic properties of the system; and it shows in particular that the exceptional case (11) can arise only in highly dissipative systems. The simplest example for the phenomenon will be given in §7.

In the case of an enumerable space  $E$  we write corresponding to (4)

$$(12) \quad p(t, x_i) = p_i(t), \quad \Pi(t, x_i, x_k) = \Pi_{ik}(t).$$

Equations (8) and (9) are then equivalent with

$$(13) \quad \frac{\partial P_{ik}(\tau, t)}{\partial t} = -p_k(t)P_{ik}(\tau, t) + \sum_j p_j(t)\Pi_{jk}(t)P_{ij}(\tau, t)$$

and

$$(14) \quad \frac{\partial P_{ik}(\tau, t)}{\partial \tau} = p_i(\tau) \left\{ P_{ik}(\tau, t) - \sum_j \Pi_{ij}(\tau)P_{jk}(\tau, t) \right\}.$$

In this case the condition (7) is obviously only a regularity restriction, and there exists only the type of purely discontinuous processes. It follows from the results of the present paper that (7) implies the existence of  $\partial P_{ik}(\tau, t)/\partial t$  for almost all  $t$ , and hence also the convergence of the sum in (13) for almost all  $t$ . However, this sum may diverge for special values of  $t$ . It is easy to impose on  $p_i(t)$  and  $\Pi_{ik}(t)$  further restrictions ensuring the convergence of the sum in (13) for all  $t$  (cf. §2, (23)).

Equations (13)–(14) were derived by Kolmogoroff [6] under some slight additional hypothesis on the passage to the limit in (7). The case of finitely many  $x_i$  was dealt with by several different methods: a full account of them is to be found in Fréchet's treatise [5]. In the case of infinitely many  $x_i$ , the first attempt was made by Kolmogoroff, who found a sufficient condition for the existence of a solution of (13) with the initial condition (5)<sup>(\*)</sup>. From the results of the present paper it readily follows, however, that Kolmogoroff's solution is not necessarily a probability distribution, since it is possible that

(\*) Kolmogoroff [6], §10. The usual notation is:  $-p_i(t) = A_{ii}(t)$ ,  $p_i(t)\Pi_{ik}(t) = A_{ik}(t)$ . With this notation Kolmogoroff's condition requires that, putting  $B_i^{(0)} = 1$  and  $B_i^{(n+1)} = \sum_j B_j^{(n)}|A_{jk}|$ , all  $B_i^{(n)}$  exist and that  $\sum_n B_i^{(n)}x^n/n!$  converges for some  $x > 0$  and all  $k$ .

$\sum_k P_{ik} < 1$ . On the other hand, Kolmogoroff's assumptions are rather restrictive.

The case of  $E$ 's being the real axis or any Borel set on it was dealt with by Feller [3]. Equations (8)–(9) were, however, derived from (7) under additional hypothesis, and integrated only in the case of a bounded  $p(t, x)$  (in which case (11) cannot occur). It may be pointed out that this covers also the special case of enumerable  $E$  in the case of bounded coefficient in (13)–(14)<sup>(9)</sup>. As Dubrowski [2] has shown, Feller's results and proofs can be transferred almost literally to the case of an arbitrary abstract space  $E$ <sup>(10)</sup>.

The present method of dealing with equations (8)–(9) is more general than that used loc. cit. [2, 3], but affords at the same time a considerable simplification. The same simplification can be made in the treatment of the more general integro-differential equation of parabolic type:

$$(15) \quad \frac{\partial P(\tau, x; t, \Lambda)}{\partial \tau} + a(\tau, x) \frac{\partial^2 P(\tau, x; t, \Lambda)}{\partial x^2} + b(\tau, x) \frac{\partial P(\tau, x; t, \Lambda)}{\partial x} \\ = p(\tau, x) \left\{ P(\tau, x; t, \Lambda) - \int_E P(\tau, y; t, \Lambda) \Pi(\tau, x, dE_y) \right\},$$

where  $E$  is the real axis,  $a(\tau, x) > 0$ . This equation and its adjoint describe the mixed type of a Markoff process<sup>(11)</sup>.

**2. Preliminaries.** The following assumptions on  $p(t, x)$  and  $\Pi(t, x, \Lambda)$  will be made throughout the paper:

(i)  $p(t, x)$  is finite and non-negative for all points  $x$  of  $E$  and all  $t$  of some finite or infinite interval  $T_0 < t < T_1$ . For  $x$  fixed,  $p(t, x)$  is a continuous function of  $t$ , and for  $t$  fixed it is measurable with respect to  $\mathfrak{B}$ .

(ii)  $\Pi(t, x, \Lambda)$  is defined for  $T_0 < t < T_1$ , for all  $x \in E$  and all sets  $\Lambda \in \mathfrak{B}$ . For fixed  $x, \Lambda$  it is a continuous function of  $t$ ; for fixed  $t, \Lambda$  it is measurable with respect to  $\mathfrak{B}$ , and for fixed  $t, x$  it is a non-negative completely additive func-

<sup>(9)</sup> It suffices namely to interpret the points  $x_k$  as integers. It was with a view of this case that  $p(t, x)$  was, in [3], supposed only to be measurable with respect to  $x$ . The point was not, however, mentioned explicitly and seems to have been generally overlooked.

<sup>(10)</sup> Added in proof: In a recent paper [9] (which became accessible to the author only after the present paper was submitted for publication), W. Doeblin investigated essentially the same class of stochastic processes with which we are concerned here. It may be remarked that Doeblin's methods as well as his results are different from ours. He proceeds by a direct and careful analysis of the stochastic movement itself, and arrives at a characterization of the process by means of two functions  $U(\tau, t, x)$  and  $V(\tau, x; t, \Lambda)$  which may, roughly, be described, respectively, as the probability that the moving point  $X(t)$  will remain in its initial position  $x$  during  $(\tau, t)$ , and the compound probability that it will undergo a change such that the first jump takes it into the set  $\Lambda$ . These functions must satisfy the functional equations  $U(\tau, t, x) = U(\tau, s, x)U(s, t, x)$  and  $V(\tau, x; t, \Lambda) = V(\tau, x; s, \Lambda) + U(\tau, s, x)V(s, x; t, \Lambda)$  for  $\tau < s < t$ . It is shown that except for these equations and some trivial additional restrictions the functions  $U$  and  $V$  can be prescribed arbitrarily. The occurrence of the exceptional case (11) is ruled out by a uniformity condition.

<sup>(11)</sup> For the definition see §2 and for the integration of (15), §5 of Feller [3].

tion of sets  $\Lambda \in \mathfrak{B}$  with

$$(16) \quad \Pi(t, x, E) = 1.$$

Finally, for the set  $\Lambda = x$  we suppose that

$$(17) \quad \Pi(t, x, x) = 0.$$

Throughout this paper the parameters  $t$  and  $\tau$  are restricted so that

$$T_0 < \tau < t < T_1$$

where  $(T_0, T_1)$  is the interval specified above.  $x, y, z$  will denote points of  $E$ . Any function of points will be supposed, or is easily seen to be, measurable with respect to  $\mathfrak{B}$ . A set  $\Lambda \in \mathfrak{B}$  will be called bounded if  $p(t, x)$  is uniformly bounded for all  $t$  and  $x \in \Lambda$ . In particular, we shall write

$$(18) \quad \Lambda_a = E \{ p(t, x) < a \},$$

where  $a > 0$ . By (i) obviously  $\Lambda_a \in \mathfrak{B}$  and  $\Lambda_a \uparrow E$  as  $a \uparrow \infty$ . Any finite set is bounded, and in the case of an enumerable  $E$  it is more convenient to consider finite sets instead of bounded. A similar remark applies if  $E$  is equipped with a metric.

By (i) and (ii) integrals of the type  $J(t, x, \Lambda) = \int_{\Lambda} p(t, y) \Pi(t, x, dE_y)$  have a meaning; if, in particular,  $\Lambda$  is a bounded set,  $J(t, x, \Lambda)$  is for fixed  $x$  a continuous function of  $t$ , and for fixed  $t$  a function of  $x$  which is measurable with respect to  $\mathfrak{B}$ . Now any set  $\Lambda \in \mathfrak{B}$  is the limit of an increasing sequence of bounded sets, and hence any function of the type  $J(t, x, \Lambda)$  is the limit of a monotonic sequence of functions which are, for fixed other arguments, continuous with respect to  $t$  and measurable with respect to  $\mathfrak{B}$ . This remark applies to all integrals which will be used in the sequel, and enables us in particular to use repeated integrals. We shall also frequently have to interchange the order of integration. To legitimate this procedure once for all the following may be remarked.

Only two different types of inversions will be used. Sometimes both integrations will be with respect to time-parameters: in such cases the elementary theory of repeated integrals will suffice to justify the change in the order of integration. In all other cases the inversion will be based on the following

LEMMA<sup>(12)</sup>. Let  $E^{(1)}$  and  $E^{(2)}$  be two spaces, and let  $\mathfrak{B}^{(i)}$  be a Borel field of subsets of  $E^{(i)}$ ,  $i = 1, 2$ . Denote by  $x^{(i)}$  a point varying in  $E^{(i)}$ , and by  $\Lambda^{(i)}$  a set belonging to  $\mathfrak{B}^{(i)}$ . Let  $f(x^{(1)})$  and  $g(x^{(2)})$  be two non-negative and bounded functions, measurable with respect to  $\mathfrak{B}^{(1)}$  and to  $\mathfrak{B}^{(2)}$ , respectively. Let  $F(\Lambda^{(1)})$  be a completely additive function of sets  $\Lambda^{(1)} \in \mathfrak{B}^{(1)}$ , with  $0 \leq F(\Lambda^{(1)}) \leq 1$ . Finally, let  $G(x^{(1)}, \Lambda^{(2)})$  be defined for all  $x^{(1)} \in E^{(1)}$  and  $\Lambda^{(2)} \in \mathfrak{B}^{(2)}$  so that it is, for fixed  $x^{(1)}$ ,

<sup>(12)</sup> Added in proof: A similar theorem for the case of the real axis was announced by R. H. Cameron and W. T. Martin, but is not yet published; see the abstract presented to the American Mathematical Society, 46-3-162.

a completely additive function of sets  $\Lambda^{(2)}$ , and for fixed  $\Lambda^{(2)}$  measurable with respect to  $\mathfrak{B}^{(1)}$ , and so that for all values of the arguments  $0 \leq G(x^{(1)}, \Lambda^{(2)}) \leq 1$ . Then for any two fixed sets  $\Gamma^{(1)} \in \mathfrak{B}^{(1)}$  and  $\Gamma^{(2)} \in \mathfrak{B}^{(2)}$

$$(19) \quad \int_{\Gamma^{(1)}} f(x^{(1)}) F(dE_x^{(1)}) \int_{\Gamma^{(2)}} g(x^{(2)}) G(x^{(1)}, dE_x^{(2)}) \\ = \int_{\Gamma^{(2)}} g(x^{(2)}) \int_{\Gamma^{(1)}} f(x^{(1)}) G(x^{(1)}, dE_x^{(2)}) F(dE_x^{(1)}).$$

Before proving this theorem let us remark that it is much simpler than Fubini's theorem, but is not contained in it. In our applications either both spaces  $E^{(i)}$  will coincide with  $E$ , or else  $E^{(1)}$  will be the real time-axis and  $E^{(2)}$  the space  $E$ . It is clear that

$$\int_{\Gamma^{(2)}} g(x^{(2)}) G(x^{(1)}, dE_x^{(2)})$$

is, as a function of  $x^{(1)}$ , measurable with respect to  $\mathfrak{B}^{(1)}$ , since it is the limit of measurable functions. Similarly

$$\int_{\Gamma^{(1)}} f(x^{(1)}) G(x^{(1)}, \Lambda^{(2)}) F(dE_x^{(1)})$$

is a completely additive function of sets  $\Lambda^{(2)} \in \mathfrak{B}^{(2)}$ . Thus both sides in (19) have a meaning.

The lemma is easily proved by a decomposition  $\Gamma^{(2)} = \sum_n \Gamma_n^{(2)}$  where  $\Gamma_n^{(2)}$  is the set of all points  $x^{(2)}$  where  $(n-1)\epsilon \leq g(x^{(2)}) < n\epsilon$ . Since  $g(x^{(2)})$  is bounded, only finitely many  $\Gamma_n^{(2)}$  are not empty. Hence

$$\begin{aligned} & \int_{\Gamma^{(2)}} g(x^{(2)}) \int_{\Gamma^{(1)}} f(x^{(1)}) G(x^{(1)}, dE_x^{(2)}) F(dE_x^{(1)}) \\ & \leq \sum_n n\epsilon \int_{\Gamma^{(1)}} f(x^{(1)}) G(x^{(1)}, \Gamma_n^{(2)}) F(dE_x^{(1)}) \\ & \leq \int_{\Gamma^{(1)}} f(x^{(1)}) F(dE_x^{(1)}) \sum_n \left\{ \epsilon G(x^{(1)}, \Gamma_n^{(2)}) + \int_{\Gamma_n^{(2)}} g(x^{(2)}) G(x^{(1)}, dE_x^{(2)}) \right\} \\ & \leq \int_{\Gamma^{(1)}} f(x^{(1)}) F(dE_x^{(1)}) \int_{\Gamma^{(2)}} g(x^{(2)}) G(x^{(1)}, dE_x^{(2)}) + \epsilon \int_{\Gamma^{(1)}} f(x^{(1)}) F(dE_x^{(1)}), \end{aligned}$$

and the last integral is bounded. This proves (19) with the sign  $\geq$  instead of the equality. In the same way, however, we get also the opposite limitation, and this accomplishes the proof.

A word has still to be said about the derivation of the equations (8) and (9) and the relations between them, though this is by no means necessary for

the understanding of the following existence theorems. Accordingly, the reader can pass over directly to §3.

Equation (8) is more natural than (9) since, roughly speaking, (9) describes the process in its dependence on the initial values. (8) leads also to a representation of  $P(\tau, x; t, \Lambda)$  which is in most cases more useful than the representation deduced from (9). The later equation is, nevertheless, simpler than (8) since the integrals in (9) converge for all sets  $\Lambda \in \mathfrak{B}$  and a derivative  $\partial P/\partial \tau$  exists for all  $\tau$  whereas the integrals in (8) will, in general, converge only for bounded sets and almost all  $t$ , so that also  $\partial P/\partial t$  exists only for bounded sets and almost all  $t$ .

In previous papers<sup>(13)</sup> the equations (8) and (9) were derived under the assumption that the passage to the limit in (7) takes place uniformly with respect to  $x$ . For a general theory, however, such an assumption is not only an unnecessary restriction, but is also dangerous since it can be shown by examples that it is not realized for the actual solutions<sup>(14)</sup>.

To deduce (9) we observe that by (3) we have for  $\Delta\tau > 0$

$$P(\tau - \Delta\tau, x; t, \Lambda) = \int_E P(\tau - \Delta\tau, x; \tau, dE_y) P(\tau, y; t, \Lambda)$$

or, splitting the space of integration into  $x$  and  $E - x$ ,

$$\begin{aligned} & - \frac{1}{\Delta\tau} \{ P(\tau - \Delta\tau, x; t, \Lambda) - P(\tau, x; t, \Lambda) \} \\ (20) \quad & = P(\tau, x; t, \Lambda) \frac{P(\tau - \Delta\tau, x; \tau, x) - 1}{-\Delta\tau} \\ & \quad - \frac{1}{\Delta\tau} \int_{E-x} P(\tau, y; t, \Lambda) P(\tau - \Delta\tau, x; \tau, dE_y). \end{aligned}$$

Now by (7) and (17)

$$- \frac{1}{\Delta\tau} \{ P(\tau - \Delta\tau, x; \tau, x) - 1 \} \rightarrow p(\tau, x),$$

and using (16) and (17) it is seen that also

$$\frac{1}{\Delta\tau} P(\tau - \Delta\tau, x; \tau, E - x) \rightarrow p(\tau, x).$$

Hence, for fixed  $\tau, x$ , the ratio  $P(\tau - \Delta\tau, x; \tau, \Lambda)/\Delta\tau$  is uniformly bounded for all sets  $\Lambda$  not containing  $x$ , and by (7) this quantity tends to  $p(\tau, x)\Pi(\tau, x, \Lambda)$ .

<sup>(13)</sup> Feller [3], Dubrovski [2].

<sup>(14)</sup> It may be remarked that the occurrence of solutions satisfying (11) has nothing whatsoever to do with the nonuniformity of the passage to the limit in (7) (or with the circumstance that the derivatives of  $P(\tau, x; t, \Lambda)$  are not bounded).

The right-hand member of (20) is thus seen to tend to the limit given by (9), and it follows from (20) that a left-hand derivative  $\partial P(\tau, x; t, \Lambda)/\partial \tau$  exists for all  $\tau, x, t, \Lambda$  and that with this derivative (9) holds. The actual (and unique) solution of (9) will show that this left-hand derivative actually is the derivative in the usual sense.

It seems impossible to give a strict proof also for (8) in an equally simple way. One can easily render (8) plausible by writing, according to (3),

$$(21) \quad \frac{1}{\Delta t} \{ P(\tau, x; t + \Delta t, \Lambda) - P(\tau, x; t, \Lambda) \} \\ = \int_E P(\tau, x; t, dE_y) \{ P(t, y; t + \Delta t, \Lambda) - \delta(y, \Lambda) \} / \Delta t,$$

and going formally to the limit applying (7). In a strict sense, however, one gets by this procedure only a partial result. Denote namely by  $\mathfrak{B}'$  the class of sets such that  $\Lambda \in \mathfrak{B}'$  if, and only if,  $\Lambda$  is bounded and there is a constant  $a > 0$  such that

$$(22) \quad \frac{1 - P(t, x; t + \Delta t, x)}{\Delta t} < a, \quad \frac{P(t, x; t + \Delta t, E - x)}{\Delta t} < a$$

for all  $t$ , all  $x \in \Lambda$ , and all  $\Delta t > 0$ . Denoting then by  $D_t$  the upper right-hand derivative with respect to  $t$ , it follows easily from (21) and (22) that for all sets  $\Lambda \in \mathfrak{B}'$

$$D_t P(\tau, x; t, \Lambda) \geq - \int_{\Lambda} p(t, y) P(\tau, x; t, dE_y) \\ + \int_E p(t, y) \Pi(t, y, \Lambda) P(\tau, x; t, dE_y).$$

Now here the first integral converges, since  $\Lambda$  is a bounded set. Thus for fixed  $\Lambda \in \mathfrak{B}'$  and for fixed  $x$ ,  $D_t P(\tau, x; t, \Lambda)$  is uniformly bounded from below; and  $D_t P(\tau, x; t, \Lambda) = \infty$  for all values of  $t$  for which the second integral diverges. Since  $0 \leq P(\tau, x; t, \Lambda) \leq 1$  it follows that, for  $\Lambda \in \mathfrak{B}'$ , the second integral must converge for almost all  $t$ , that it is to say, that a finite right-hand derivative  $\partial P(\tau, x; t, \Lambda)/\partial t$  exists for all  $\tau, \Lambda \in \mathfrak{B}'$ , and almost all  $t$ , and furthermore that with this derivative

$$\frac{\partial P(\tau, x; t, \Lambda)}{\partial t} \geq - \int_{\Lambda} p(t, y) P(\tau, x; t, dE_y) + \int_E p(t, y) \Pi(t, y, \Lambda) P(\tau, x; t, dE_y).$$

Actually the sign of equality in (8) holds not only for all  $\Lambda \in \mathfrak{B}'$  but even for all bounded sets and almost all  $t$ . For the sake of simplicity we prefer, however, to prove this assertion in an indirect way:



We shall namely prove that *there is* (under the assumptions (i)-(ii) on  $p(t, x)$  and  $\Pi(t, x, \Lambda)$ ) *one and only one function*  $P(\tau, x; t, \Lambda)$  *satisfying* (9) *with the initial condition* (2) *and which is, for fixed*  $\tau, x, t$  *a completely additive function of sets*  $\Lambda \in \mathfrak{B}$  *with*  $0 \leq P(\tau, x; t, \Lambda) \leq 1$ . *For bounded sets*  $\Lambda$  *this function will be shown to be an absolutely continuous function of*  $t$ , *satisfying* (8) *for almost all*  $t$ .

Moreover, it will be shown that with *this solution* (3) *also holds*. This gives a uniqueness theorem for our general problem, but an existence theorem will be given essentially only for uniformly bounded  $p(t, x)$  (cf. Theorem 6), since sometimes instead of (1) only (11) holds.

This result shows in particular that we may use, instead of the class  $\mathfrak{B}'$  considered above, the class  $\mathfrak{B}_1$  of all bounded sets. This by itself does not imply that (22) holds for any bounded set and some suitable  $a$ . It may, however, be remarked that this is actually the case, as is readily seen from the representation of the solution given below.

It may still be pointed out that it can be shown by examples that not even for bounded sets  $\Lambda$  does the derivative  $\partial P(\tau, x; t, \Lambda)/\partial t$  need to exist for all  $t$ . It is, however, easy to make additional assumptions on  $p(t, x)$  which assure the existence of  $\partial P(\tau, x; t, \Lambda)/\partial t$  for all bounded sets and all  $t$ . Such a hypothesis is, for instance, that

$$(23) \quad \frac{p(t_1, x)}{1 + p(t_2, x)} < M$$

uniformly for all values  $x, t_1, t_2$ . This hypothesis is in particular fulfilled in the case of temporally homogeneous processes.

3. *Solution of* (8). We shall define a new completely additive function of sets  $\Lambda \in \mathfrak{B}$  by

$$(24) \quad \Pi^*(\tau, x; t, \Lambda) = \int_{\Lambda} \exp \left\{ - \int_{\tau}^t p(s, y) ds \right\} \Pi(\tau, x, dE_y).$$

Obviously  $0 \leq \Pi^*(\tau, x; t, \Lambda) \leq 1$ ; furthermore for any bounded set  $\Lambda$

$$(25) \quad \begin{aligned} \frac{\partial \Pi^*(\tau, x; t, \Lambda)}{\partial t} &= - \int_{\Lambda} p(t, y) \exp \left\{ - \int_{\tau}^t p(s, y) ds \right\} \Pi(\tau, x, dE_y) \\ &= - \int_{\Lambda} p(t, y) \Pi^*(\tau, x; t, dE_y). \end{aligned}$$

THEOREM 1.  $P_{ut}^{(15)}$

$$(26) \quad P^{(0)}(\tau, x; t, \Lambda) = \delta(x, \Lambda) \exp \left\{ - \int_{\tau}^t p(s, x) ds \right\},$$

(15)  $\delta(x, \Lambda)$  was defined by (2).

and for  $n \geq 1$

$$(27) \quad P^{(n)}(\tau, x; t, \Lambda) = \int_{\tau}^t d\sigma \int_E p(\sigma, y) \Pi^*(\sigma, y; t, \Lambda) P^{(n-1)}(\tau, x; \sigma, dE_y).$$

Let

$$(28) \quad P(\tau, x; t, \Lambda) = \sum_{n=0}^{\infty} P^{(n)}(\tau, x; t, \Lambda);$$

(i) the function  $P(\tau, x; t, \Lambda)$  is for fixed  $\tau, t, x \in E$  a completely additive function of sets  $\Lambda \in \mathfrak{B}$  with  $0 \leq P(\tau, x; t, \Lambda) \leq 1$ ; (ii)  $P(\tau, x; t, \Lambda)$  is for fixed  $\tau, x, \Lambda$  an absolutely continuous function of  $t$ ; for any bounded  $\Lambda$  and almost all  $t$  the derivative  $\partial P / \partial t$  is finite and satisfies (8) with the initial condition (2).

**Remark.** It will be seen that for any bounded  $\Lambda$  and almost all  $t$

$$(29) \quad \begin{aligned} \frac{\partial P^{(0)}(\tau, x; t, \Lambda)}{\partial t} &= -p(t, x) P^{(0)}(\tau, x; t, \Lambda), \\ \frac{\partial P^{(n)}(\tau, x; t, \Lambda)}{\partial t} &= - \int_{\Lambda} p(t, y) P^{(n)}(\tau, x; t, dE_y) \\ &\quad + \int_E p(t, y) \Pi(t, y, \Lambda) P^{(n-1)}(\tau, x; t, dE_y); \end{aligned}$$

the integrals on the right side converging for almost all  $t$ . These equations are in close analogy with (8), and afford the interpretation of  $P^{(n)}(\tau, x; t, \Lambda)$  as the compound probability that during  $(\tau, t)$  the state  $X$  will change by exactly  $n$  jumps and that  $X(t) \in \Lambda$ , if it is known that  $X(\tau) = x$ .

In the special case of an enumerable  $E$  it is, of course, sufficient to determine the quantities  $P_{ik}^{(n)}(\tau, t)$ . For these, (29) reduces to the ordinary differential equations

$$(30) \quad \begin{aligned} \frac{\partial P_{ik}^{(0)}(\tau, t)}{\partial t} &= -p_k(t) P_{ik}^{(0)}(\tau, t), \\ \frac{\partial P_{ik}^{(n)}(\tau, t)}{\partial t} &= -p_k(t) P_{ik}^{(n)}(\tau, t) + \sum_j p_j(t) \Pi_{jk}(t) P_{ij}^{(n-1)}(\tau, t), \end{aligned}$$

and (26)–(27) to

$$P_{ij}^{(n)}(\tau, t) = \sum_j \int_{\tau}^t \exp \left\{ - \int_{\tau}^t p_k(s) ds \right\} p_j(\sigma) \Pi_{jk}(\sigma) P_{ij}^{(n-1)}(\tau, \sigma) d\sigma.$$

If the  $p_j(\sigma)$  are not subjected to a further restriction analogous to (23), the derivative  $\partial P_{ik}^{(n)}(\tau, t) / \partial t$  will exist only for almost all  $t$ .

**Proof.** Suppose, by induction, that  $P^{(n)}(\tau, x; t, \Lambda)$  exists for some fixed  $n \geq 0$ , and all values of the arguments, and that it is a completely additive function of sets  $\Lambda \in \mathfrak{B}$  with  $0 \leq P^{(n)}(\tau, x; t, \Lambda) \leq 1$ ; furthermore that

$$(31) \quad L^{(n)}(\tau, x, t) = \int_{\tau}^t d\sigma \int_E p(\sigma, y) P^{(n)}(\tau, x; \sigma, dE_y)$$

is finite. This is certainly true for  $n=0$  and

$$(32) \quad P^{(0)}(\tau, x; t, E) + L^{(0)}(\tau, x, t) = 1.$$

It follows then from (27) that also  $P^{(n+1)}(\tau, x; t, \Lambda)$  exists and  $0 \leq P^{(n+1)}(\tau, x; t, \Lambda) \leq L^{(n)}(\tau, x, t)$ . For any bounded set  $\Lambda$ , therefore, we get from (25) and (27)

$$\begin{aligned} \int_{\Lambda} p(t, y) P^{(n+1)}(\tau, x; t, dE_y) \\ = - \int_{\tau}^t d\sigma \int_E p(\sigma, y) \frac{\partial \Pi^*(\sigma, y; t, \Lambda)}{\partial t} P^{(n)}(\tau, x; \sigma, dE_y); \end{aligned}$$

the left-hand member is obviously a continuous function of  $t$ , and we get

$$\begin{aligned} \int_{\tau}^t d\sigma_1 \int_{\Lambda} p(\sigma_1, y) P^{(n+1)}(\tau, x; \sigma_1, dE_y) \\ = - \int_{\tau}^t d\sigma_1 \int_{\tau}^{\sigma_1} d\sigma \int_E p(\sigma, y) \frac{\partial \Pi^*(\sigma, y; \sigma_1, \Lambda)}{\partial \sigma_1} P^{(n)}(\tau, x; \sigma, dE_y); \end{aligned}$$

inverting the order of integration and observing that  $\Pi^*(\sigma, y; \sigma, \Lambda) = \Pi(\sigma, y, \Lambda)$ , we get

$$\begin{aligned} \int_{\tau}^t d\sigma_1 \int_{\Lambda} p(\sigma_1, y) P^{(n+1)}(\tau, x; \sigma_1, dE_y) \\ = \int_{\tau}^t d\sigma \int_E p(\sigma, y) \Pi(\sigma, y, \Lambda) P^{(n)}(\tau, x; \sigma, dE_y) \\ - \int_{\tau}^t d\sigma \int_E p(\sigma, y) \Pi^*(\sigma, y; t, \Lambda) P^{(n)}(\tau, x; \sigma, dE_y), \end{aligned}$$

or by (27) finally

$$\begin{aligned} P^{(n+1)}(\tau, x; t, \Lambda) + \int_{\tau}^t d\sigma \int_{\Lambda} p(\sigma, y) P^{(n+1)}(\tau, x; \sigma, dE_y) \\ (33) \quad = \int_{\tau}^t d\sigma \int_E p(\sigma, y) \Pi(\sigma, y, \Lambda) P^{(n)}(\tau, x; \sigma, dE_y). \end{aligned}$$

This is, essentially, the relation (29) of the remark following Theorem 1. Equation (33) holds for any bounded set  $\Lambda$ . We apply (33) in particular to

$\Lambda = \Lambda_a$  (see (18)) and let  $a \uparrow \infty$ . Since  $\Pi(\sigma, y, \Lambda_a) \leq 1$  the right-hand member is bounded by  $L^{(n)}(\tau, x, t)$  (cf. (31)). Hence we get

$$(34) \quad P^{(n+1)}(\tau, x; t, E) + L^{(n+1)}(\tau, x, t) = L^{(n)}(\tau, x, t).$$

It is thus seen that both  $P^{(n+1)}(\tau, x; t, \Lambda)$  and  $L^{(n+1)}(\tau, x, t)$  exist. Moreover, since  $P^{(n+1)}(\tau, x; t, \Lambda) \geq 0$ , we have  $L^{(n+1)}(\tau, x, t) \leq L^{(n)}(\tau, x, t)$ . Thus, for all  $n \geq 0$ ,

$$(35) \quad 1 \geq L^{(0)}(\tau, x, t) \geq L^{(1)}(\tau, x, t) \geq \dots \geq L^{(n)}(\tau, x, t) \rightarrow L(\tau, x, t).$$

By (34) and (35) also  $P^{(n+1)}(\tau, x; t, \Lambda) \leq 1$ , and thus the assumptions of the inductive argument hold for all  $n$ .

It may be remarked that it can be shown by examples that the integrand of (31),  $\int_E p(\sigma, y) P^{(n)}(\tau, x; \sigma, dE_y)$ , sometimes diverges for some values of  $\sigma$ . The proof shows, however, that it converges for almost all  $\sigma$ , and it is readily seen that it converges for all  $\sigma$  if (23) holds.

Now we get from (34) and (32)

$$(36) \quad \sum_{n=0}^N P^{(n)}(\tau, x; t, E) = 1 - L^{(N)}(\tau, x, t),$$

and thus  $0 \leq P(\tau, x; t, \Lambda) \leq 1$ .

Hence we readily deduce from (33) for any bounded set  $\Lambda \in \mathfrak{B}$

$$(37) \quad \begin{aligned} P(\tau, x; t, \Lambda) + \int_{\tau}^t d\sigma \int_{\Lambda} p(\sigma, y) P(\tau, x; \sigma, dE_y) \\ = \int_{\tau}^t d\sigma \int_E p(\sigma, y) \Pi(\sigma, y, \Lambda) P(\tau, x; \sigma, dE_y); \end{aligned}$$

this proves (8) for almost all  $t$ . If (23) holds, (37) can obviously be differentiated for all  $t$  and (8) holds for any bounded set and all  $t$ .

From (35) and (37) we get also the following

**COROLLARY.** *The necessary and sufficient condition that  $P(\tau, x; t, E) = 1$  for all  $t$  is that*

$$(38) \quad L(\tau, x, t) = \lim_{N \rightarrow \infty} L^{(N)}(\tau, x, t) = 0$$

for all  $t$ .

Incidentally, it is quite obvious that (7) holds at least for all bounded sets  $\Lambda$ , since by (26) and (27)

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \{ \delta(x, \Lambda) - P^{(0)}(\tau, x; \tau + h, \Lambda) \} &= p(\tau, x) \delta(x, \Lambda), \\ \lim_{h \rightarrow 0} \frac{1}{h} P^{(1)}(\tau, x; \tau + h, \Lambda) &= p(\tau, x) \Pi(\tau, x, \Lambda) \end{aligned}$$

and by (10) and (16)

$$\begin{aligned} \limsup_{h \rightarrow +0} \frac{1}{h} \sum_{n=2}^{\infty} P^{(n)}(\tau, x; \tau + h, E) \\ \leq \lim_{h \rightarrow +0} \frac{1}{h} \{1 - P^{(0)}(\tau, x; \tau + h, E) - P^{(1)}(\tau, x; \tau + h, E)\} \\ = 0. \end{aligned}$$

That (7) holds for any  $\Lambda \in \mathfrak{B}$  will be proved in §5.

4. **Solution of (9).** We now prove the following theorem.

**THEOREM 2.** Put

$$(39) \quad Q^{(0)}(\tau, x; t, \Lambda) = \delta(x, \Lambda) \exp \left\{ - \int_{\tau}^t p(s, x) ds \right\},$$

and for  $n \geq 1$

$$\begin{aligned} (40) \quad Q^{(n)}(\tau, x; t, \Lambda) = \int_{\tau}^t p(\sigma, x) \\ \exp \left\{ - \int_{\tau}^{\sigma} p(s, x) ds \right\} d\sigma \int_{\mathfrak{B}} Q^{(n-1)}(\sigma, y; t, \Lambda) \Pi(\sigma, x, dE_y). \end{aligned}$$

Then<sup>(10)</sup>, for any fixed  $\tau, x, t$ ,

$$(41) \quad P(\tau, x; t, \Lambda) = \sum_{n=0}^{\infty} Q^{(n)}(\tau, x; t, \Lambda)$$

is a completely additive function of sets  $\Lambda \in \mathfrak{B}$  and  $0 \leq P(\tau, x; t, \Lambda) \leq 1$ . Furthermore  $P(\tau, x; t, \Lambda)$  is a solution of (9) with the initial values given by (2).

**Remark.** Obviously the  $Q^{(n)}(\tau, x; t, \Lambda)$  are solutions of the equations

$$\begin{aligned} \frac{\partial Q^{(0)}(\tau, x; t, \Lambda)}{\partial \tau} &= p(\tau, x) Q^{(0)}(\tau, x; t, \Lambda), \\ (42) \quad \frac{\partial Q^{(n)}(\tau, x; t, \Lambda)}{\partial \tau} &= p(\tau, x) \left\{ Q^{(n)}(\tau, x; t, \Lambda) \right. \\ &\quad \left. - \int_{\mathfrak{B}} Q^{(n-1)}(\tau, y; t, \Lambda) \Pi(\tau, x, dE_y) \right\}, \end{aligned}$$

which can be treated as ordinary differential equations.

**Proof.** Put

$$S^{(n)}(\tau, x; t, \Lambda) = \sum_{k=0}^n Q^{(k)}(\tau, x; t, \Lambda).$$

<sup>(10)</sup> It will be seen (Theorem 4) that the functions defined by (41) and (28) are actually identical.

Then  $0 \leq S^{(0)}(\tau, x; t, \Lambda) \leq 1$ . Let us suppose that  $0 \leq S^{(0)} \leq S^{(1)} \leq \dots \leq S^{(n-1)} \leq 1$ . Then by (39) and (40)

$$\begin{aligned}
 S^{(n)}(\tau, x; t, \Lambda) &= \exp \left\{ - \int_{\tau}^t p(s, x) ds \right\} \left\{ \delta(x, \Lambda) + \int_{\tau}^t p(\sigma, x) \right. \\
 (43) \quad &\exp \left\{ \int_{\sigma}^t p(s, x) ds \right\} d\sigma \int_{\mathcal{B}} S^{(n-1)}(\sigma, y; t, \Lambda) \Pi(\sigma, x, dE_y) \Big\} \\
 &\leq \exp \left\{ - \int_{\tau}^t p(s, x) ds \right\} \left\{ 1 \right. \\
 &\quad \left. + \int_{\tau}^t p(\sigma, x) \exp \left\{ \int_{\sigma}^t p(s, x) ds \right\} d\sigma \right\} = 1.
 \end{aligned}$$

On the other hand obviously  $S^{(n)}(\tau, x; t, \Lambda) \geq S^{(n-1)}(\tau, x; t, \Lambda)$ . Hence  $S^{(n)}(\tau, x; t, \Lambda) \uparrow P(\tau, x; t, \Lambda) \leq 1$ . That  $P(\tau, x; t, \Lambda)$  is a solution of (9) follows immediately from (42), and also the initial condition (2) is obviously satisfied.

**THEOREM 3** (Uniqueness theorem<sup>(17)</sup>). Consider some fixed  $t$  and a function  $P^*(\tau, x; t, \Lambda)$  which (i) for fixed  $\tau, x, t$  is a completely additive function of sets  $\Lambda \in \mathcal{B}$  with  $0 \leq P^*(\tau, x; t, \Lambda) \leq 1$ ; (ii) for fixed  $x, t, \Lambda$  is an absolutely continuous function of  $\tau$  satisfying for almost all  $\tau$  the equation (9) with the initial value (2) as  $\tau \rightarrow t - 0$ . Then  $P^*(\tau, x; t, \Lambda) = P(\tau, x; t, \Lambda)$ , where  $P(\tau, x; t, \Lambda)$  is the function defined by Theorem 2.

**Proof.** (i) We first show that

$$(44) \quad P^*(\tau, x; t, \Lambda) \geq P(\tau, x; t, \Lambda);$$

this remains true also if the assumption  $P^*(\tau, x; t, \Lambda) \leq 1$  be replaced by the weaker one that  $P^*(\tau, x; t, \Lambda)$  is uniformly bounded. In fact, treating (9) as an ordinary differential equation, we get by (2)

$$\begin{aligned}
 P^*(\tau, x; t, \Lambda) &= \exp \left\{ - \int_{\tau}^t p(s, x) ds \right\} \left\{ \delta(x, \Lambda) + \int_{\tau}^t p(\sigma, x) \right. \\
 (45) \quad &\exp \left\{ - \int_{\sigma}^t p(s, x) ds \right\} d\sigma \int_{\mathcal{B}} P^*(\sigma, y; t, \Lambda) \Pi(\sigma, x, dE_y) \Big\}.
 \end{aligned}$$

Since the last term is non-negative, we see by comparison of (45) with (39) that  $P^*(\tau, x; t, \Lambda) \geq Q^{(0)}(\tau, x; t, \Lambda) = S^{(0)}(\tau, x; t, \Lambda)$ . Comparing, then, (45) with (43) it is readily seen that  $P^*(\tau, x; t, \Lambda) \geq S^{(n)}(\tau, x; t, \Lambda)$  for any  $n$ , which proves (44).

(ii) Put

<sup>(17)</sup> Cf. footnote 7.



$$(46) \quad D(\tau, x; t, \Lambda) = P^*(\tau, x; t, \Lambda) - P(\tau, x; t, \Lambda).$$

By (44),  $D(\tau, x; t, \Lambda)$  is a completely additive function of sets, and  $0 \leq D(\tau, x; t, \Lambda) \leq 1$ . Now, assuming that  $D(\tau, x; t, E)$  takes on the value  $\alpha > 0$  somewhere, denote by  $\tau_0$  the least upper bound of all  $\tau$  for which  $D(\tau, x; t, E) \geq \alpha$ , so that

$$(47) \quad D(\tau, x; t, E) < D(\tau_0, x; t, E) = \alpha \quad \text{for } \tau_0 < \tau < t.$$

Now  $(1/\alpha)D(\tau, x; t, \Lambda)$  is a solution of (9), which vanishes as  $\tau \rightarrow t - 0$ . Hence

$$(48) \quad \frac{1}{\alpha} D(\tau, x; t, E) = \int_{\tau}^t p(\sigma, x) \exp \left\{ - \int_{\tau}^{\sigma} p(s, x) ds \right\} d\sigma \int_E \frac{1}{\alpha} D(\sigma, y; t, \Lambda) \Pi(\sigma, x, dE_y).$$

Combining (47) and (48), we get

$$\begin{aligned} 1 &= \frac{1}{\alpha} D(\tau_0, x; t, E) \leq \int_{\tau_0}^t p(\sigma, x) \exp \left\{ - \int_{\tau_0}^{\sigma} p(s, x) ds \right\} d\sigma \\ &= 1 - \exp \left\{ - \int_{\tau_0}^t p(s, x) ds \right\} < 1. \end{aligned}$$

Thus the assumption  $D(\tau, x; t, E) = \alpha > 0$  leads to a contradiction. Hence, by (44) and (46),  $D(\tau, x; t, \Lambda) = 0$  for all sets  $\Lambda$ , and this accomplishes the proof.

5. Properties of the solutions. We now prove

THEOREM 4. (i) With the functions defined by Theorems 1 and 2 one has

$$(49) \quad P^{(n)}(\tau, x; t, \Lambda) = Q^{(n)}(\tau, x; t, \Lambda)$$

identically; thus equations (28) and (41) define the same function  $P(\tau, x; t, \Lambda)$ .

(ii) This function satisfies the fundamental assumption (7).

**Proof.** (i) Put  $P^{(n+1)}(\tau, x; t, \Lambda) = AP^{(n)}(\tau, x; t, \Lambda)$ , where  $A$  is a linear operator on  $(t, \Lambda)$ ; and similarly  $Q^{(n+1)}(\tau, x; t, \Lambda) = BQ^{(n)}(\tau, x; t, \Lambda)$  where the operator  $B$  works on  $(\tau, x)$ . Using the lemma of §2, we readily see that the operators  $A$  and  $B$  are permutable.

Now obviously  $P^{(0)}(\tau, x; t, \Lambda) = Q^{(0)}(\tau, x; t, \Lambda)$  and  $P^{(1)}(\tau, x; t, \Lambda) = Q^{(1)}(\tau, x; t, \Lambda)$ . Assuming, then, (49) to be true for some  $n \geq 1$ , we get  $P^{(n+1)} = AP^{(n)} = AQ^{(n)} = ABQ^{(n-1)} = BAP^{(n-1)} = BP^{(n)} = BQ^{(n)} = Q^{(n+1)}$ .

(ii) To prove the second part we use the representation (39)-(41). Then

$$\frac{\delta(x, \Lambda) - Q^{(0)}(\tau, x; t, \Lambda)}{t - \tau} \rightarrow p(t, x)\delta(x, \Lambda),$$

obviously, as  $\tau \rightarrow t - 0$ . Moreover

$$\sum_{n=1}^{\infty} Q^{(n)}(\tau, x; t, E) \leq 1 - Q^{(0)}(\tau, x; t, E)$$

and thus

$$(50) \quad \limsup_{\tau \rightarrow t} \frac{1}{t - \tau} \sum_{n=1}^{\infty} Q^{(n)}(\tau, x; t, E) \leq p(t, x).$$

From (40) we get however

$$(51) \quad \liminf_{\tau \rightarrow t} \frac{1}{t - \tau} \sum_{n=1}^{\infty} Q^{(n)}(\tau, x; t, \Lambda) \geq \liminf_{\tau \rightarrow t} \frac{1}{t - \tau} Q^{(1)}(\tau, x; t, \Lambda) \\ = p(t, x)\Pi(t, x, \Lambda).$$

Applying now (51) both for  $\Lambda$  and  $E - \Lambda$ , we get by (50)

$$\lim_{\tau \rightarrow t} \frac{1}{t - \tau} \sum_{n=1}^{\infty} Q^{(n)}(\tau, x; t, \Lambda) = p(t, x)\Pi(t, x, \Lambda)$$

which proves the theorem.

**THEOREM 5.** For  $\tau < \lambda < t$  one has identically

$$(52) \quad Q^{(n)}(\tau, x; t, \Lambda) = \sum_{k=0}^n \int_E Q^{(k)}(\tau, x; \lambda, dE_y) Q^{(n-k)}(\lambda, y; t, \Lambda)$$

where  $Q^{(n)}(\tau, x; t, \Lambda)$  was defined by (39)–(40).

The solution  $P(\tau, x; t, \Lambda)$  of Theorems 1 and 2 satisfies the equation (3) of Chapman-Kolmogoroff<sup>(18)</sup>.

**Proof.** The second part of the theorem is an immediate consequence of the first part.

Equation (52) is trivial for  $n=0$ . Assuming it to be true for some  $n \geq 0$ , we get by (39)–(40)

$$\sum_{k=0}^{n+1} \int_E Q^{(k)}(\tau, x; \lambda, dE_y) Q^{(n+1-k)}(\lambda, y; t, \Lambda) \\ = \exp \left\{ - \int_{\tau}^{\lambda} p(s, x) ds \right\} Q^{(n+1)}(\lambda, x; t, \Lambda) \\ + \sum_{k=1}^{n+1} \int_{\tau}^{\lambda} p(\sigma, x) \exp \left\{ - \int_{\tau}^{\sigma} p(s, x) ds \right\} d\sigma \\ \cdot \int_{E_x} \Pi(\sigma, x, dE_x) \int_{E_y} Q^{(k-1)}(\sigma, z; \lambda, dE_y) Q^{(n+1-k)}(\lambda, y; t, \Lambda)$$

<sup>(18)</sup> It should be observed that Theorem 5 is valid even in cases where  $P(\tau, x; t, \Lambda)$  is not a proper probability distribution, i.e., where  $P(\tau, x; t, E) < 1$ .

$$\begin{aligned}
&= \exp \left\{ - \int_{\tau}^{\lambda} p(s, x) ds \right\} Q^{(n+1)}(\lambda, x; t, \Lambda) \\
&\quad + \int_{\tau}^{\lambda} p(\sigma, x) \exp \left\{ - \int_{\tau}^{\sigma} p(s, x) ds \right\} d\sigma \int_{\mathbf{E}} \Pi(\sigma, x, dE_{\sigma}) Q^{(n)}(\sigma, x; t, \Lambda) \\
&= \exp \left\{ - \int_{\tau}^{\lambda} p(s, x) ds \right\} Q^{(n+1)}(\lambda, x; t, \Lambda) + Q^{(n+1)}(\tau, x; t, \Lambda) \\
&\quad - \int_{\tau}^{\lambda} p(\sigma, x) \exp \left\{ - \int_{\tau}^{\sigma} p(s, x) ds \right\} d\sigma \int_{\mathbf{E}} \Pi(\sigma, x, dE_{\sigma}) Q^{(n)}(\sigma, x; t, \Lambda) \\
&= Q^{(n+1)}(\tau, x; t, \Lambda).
\end{aligned}$$

**THEOREM 6.** *If there is some  $\alpha > 1$  and a function  $\pi(t) \in L^{\alpha}$  such that uniformly*

$$p(t, x) < \pi(t),$$

*then the solution  $P(\tau, x; t, \Lambda)$  of Theorems 1 and 2 is a probability distribution, i.e. (1) holds.*

**Proof.** With the notation of the proof of Theorem 1 we have by (35) and (31) for any  $n$

$$\begin{aligned}
L(\tau, x, t) &\leq \int_{\tau}^t d\sigma \int_{\mathbf{E}} p(\sigma, y) P^{(n)}(\tau, x; \sigma, dE_{\sigma}) \leq \int_{\tau}^t \pi(\sigma) P^{(n)}(\tau, x; \sigma, E) d\sigma \\
&\leq \left\{ \int_{\tau}^t [\pi(\sigma)]^{\alpha} d\sigma \right\}^{1/\alpha} \left\{ \int_{\tau}^t P^{(n)}(\tau, x; \sigma, E)^{\alpha/(\alpha-1)} d\sigma \right\}^{(\alpha-1)/\alpha},
\end{aligned}$$

and since  $0 \leq P^{(n)}(\tau, x; \sigma, E) \leq 1$  it follows that

$$(53) \quad \int_{\tau}^t P^{(n)}(\tau, x; \sigma, E) d\sigma \geq h(\tau, t) L(\tau, x, t)^{\alpha/(\alpha-1)},$$

where  $h(\tau, t) > 0$  is independent of  $n$ . But the left-hand member in (53) is the general term of a convergent series, and therefore  $L(\tau, x, t) = 0$ . This proves the proposition in view of the corollary to Theorem 1.

**6. The temporally homogeneous process.** So far it has been shown that there is always a function  $P(\tau, x; t, \Lambda)$  satisfying all requirements of the theory except, perhaps, (1). That (1) does not necessarily hold will be shown by means of a simple example in §7. This is a surprising result and requires a better understanding of the mechanism of the process. We shall confine ourselves to the temporally homogeneous processes; but, at least as far as sufficiency is concerned, the condition of the following theorem can easily be extended to some more general cases.

We begin with some preliminary remarks and notations. In the case of

$p(t, x)$  and  $\Pi(t, x, \Lambda)$  not depending on  $t$  the solution  $P(\tau, x; t, \Lambda)$  of Theorems 1 and 2 obviously depends only on  $t - \tau$  and we write

$$P(\tau, x; t, \Lambda) = P(t - \tau, x, \Lambda).$$

Similarly we write

$$(54) \quad p(t, x) = p(x), \quad \Pi(t, x, \Lambda) = \Pi(x, \Lambda).$$

$\Pi(x, \Lambda)$  defines in the usual way an ordinary *Markoff chain*, that is to say, a sequence of probability distributions defined by

$$(55) \quad \begin{aligned} \Pi^{(0)}(x, \Lambda) &= \delta(x, \Lambda), \\ \Pi^{(n)}(x, \Lambda) &= \int_E \Pi^{(n-1)}(y, \Lambda) \Pi(x, dE_y), \end{aligned} \quad (n \geq 1).$$

Obviously this chain is closely related to our stochastic process, and in particular the ergodic properties of the original process will be regulated by the ergodic properties of the chain (55). Roughly speaking,  $\Pi^{(n)}(x, \Lambda)$  gives the conditional probability distribution of the state  $X(t)$  under the assumption that  $X(0) = x$  and that a change of state occurred during  $(0, t)$  exactly  $n$  times—the time of occurrence of these jumps being left out of account.

If  $\Lambda$  and  $\Omega$  are any two sets of  $\mathfrak{B}$ , we put

$$(56) \quad \begin{aligned} \Pi_{\Omega}^{(0)}(x, \Lambda) &= \delta(x, \Lambda \cap \Omega), \\ \Pi_{\Omega}^{(n)}(x, \Lambda) &= \int_{\Omega} \Pi_{\Omega}^{(n-1)}(x, dE_y) \Pi(y, \Lambda), \end{aligned} \quad (n \geq 1).$$

In terms of the Markoff chain (55)  $\Pi_{\Omega}^{(n)}(x, \Lambda)$  is the probability that the moving point, starting from  $x \in \Omega$  would remain in  $\Omega$  for the  $n-1$  first steps and would be taken into some point of  $\Lambda$  by the  $n$ th step. Obviously  $\Pi_{\Omega}^{(n)}(x, \Lambda) = 0$  for all sets  $\Lambda$  if  $x$  is not contained in  $\Omega$ . For fixed  $x$  and  $\Omega$  the sequence  $\Pi_{\Omega}^{(n)}(x, \Omega)$  is never increasing:  $\Pi_{\Omega}^{(n)}(x, \Omega) \downarrow \alpha$ . If  $\alpha > 0$ , there is a positive probability of never leaving the set  $\Omega$ , if we have started from the point  $x \in \Omega$ . For further application we note that for  $x \in \Omega$  we have

$$(57) \quad \sum_{k=0}^{\infty} \Pi_{\Omega}^{(k)}(x, E - \Omega) + \Pi_{\Omega}^{(\infty)}(x, \Omega) = 1.$$

Finally we introduce the notation

$$(58) \quad \Omega_+ = E_{x \in \Omega} \{p(x) > 0\},$$

that is to say,  $\Omega_+$  consists of those points  $x \in \Omega$  for which  $p(x) \neq 0$ .

**THEOREM 7.** Suppose that  $p(t, x)$  and  $\Pi(t, x, \Lambda)$  are of the form (54).

(i) In order that the solution  $P(\tau, x; t, \Lambda)$  of Theorems 1 and 2 satisfy (1) it is necessary and sufficient that whenever for some point  $x$  and some set  $\Omega \in \mathfrak{B}$  with  $\Omega = \Omega_+$  the inequality

$$(59) \quad \Pi_{\Omega}^{(n)}(x, \Omega) > \alpha > 0$$

holds for all  $n$ , then the series

$$(60) \quad \sum_{n=0}^{\infty} \int_{\Omega} \frac{1}{p(y)} \Pi_{\Omega}^{(n)}(x, dE_y)$$

diverges<sup>(19)</sup>.

(ii) In this statement the series (60) can be replaced by

$$(61) \quad \sum_{n=0}^{\infty} \int_{\Omega} \frac{1}{p(y)} \Pi^{(n)}(x, dE_y).$$

COROLLARY. In order that  $P(t, x, E) = 1$  it is necessary that for any point  $x \in E_+$  the series

$$\sum_{n=1}^{\infty} \int_{E_+} \frac{1}{p(y)} \Pi^{(n)}(x, dE_y)$$

diverges. (This condition is, however, not sufficient, as will be shown by an example in §7.)

**Proof.** Condition (ii) is stronger than (i). We have, therefore, to prove that the divergence of (61) is a sufficient, the divergence of (60) a necessary, condition.

(i) *Sufficiency.* This part of the proof will rest mainly on the representation of  $P(t, x, \Lambda)$  given by Theorem 2.

In the case of a temporally homogeneous process the function  $L^{(n)}(\tau, x, t)$  defined by (31) depends only on  $x$  and  $t - \tau$ , and we write

$$L^{(n)}(\tau, x, t) = L^{(n)}(t - \tau, x),$$

so that for  $t > 0$

$$(62) \quad L^{(n)}(t, x) = \int_0^t d\sigma \int_E p(y) P^{(n)}(\sigma, x, dE_y).$$

Now, using the notation (54), we get from (42)

$$(63) \quad P^{(0)}(t, x, \Lambda) + p(x) \int_0^t P^{(0)}(\sigma, x, \Lambda) d\sigma = \delta(x, \Lambda),$$

<sup>(19)</sup> In the sense that the series is to be considered as divergent if some of the integrals in (60) are divergent.

and for  $n \geq 1$

$$(64) \quad \begin{aligned} P^{(n)}(t, x, \Lambda) + p(x) \int_0^t P^{(n)}(\sigma, x, \Lambda) d\sigma \\ = p(x) \int_0^t d\sigma \int_E \Pi(x, dE_y) P^{(n-1)}(\sigma, y, \Lambda). \end{aligned}$$

Combining (64) with (62) we get, for  $n \geq 1$ ,

$$(65) \quad \int_E p(y) P^{(n)}(t, x, dE_y) + p(x) L^{(n)}(t, x) = p(x) \int_E L^{(n-1)}(t, y) \Pi(x, dE_y).$$

Now, for all points  $x \in E - E_+$  (that is to say, if  $p(x) = 0$ ) we have  $L^{(n)}(t, x) = 0$  for all  $n$ . Hence, using an inductive argument, it readily follows from (65) and (32) that for  $x \in E_+$  and any  $n \geq 0$

$$(66) \quad L^{(n)}(t, x) = 1 - \sum_{k=0}^n \int_{E_+} \frac{1}{p(y)} \Pi^{(k)}(x, dE_y) \int_{E_+} p(z) P^{(n-k)}(t, y, dE_z).$$

Integrating (66) we get

$$\sum_{k=0}^n \int_{E_+} \frac{1}{p(y)} L^{(n-k)}(t, y) \Pi^{(k)}(x, dE_y) = \int_0^t \{1 - L^{(n)}(\sigma, x)\} d\sigma \leq t,$$

and since by (35)  $L^{(n)}(t, y) \downarrow L(t, y)$ , it follows that for  $x \in E_+$

$$(67) \quad \sum_{k=0}^n \int_{E_+} \frac{1}{p(y)} L(t, y) \Pi^{(k)}(x, dE_y) \leq t.$$

Suppose now that there is some point  $x_0$  and some  $t$  such that  $P(t, x_0, E) < 1$ . By the corollary to Theorem 1 this implies that

$$(68) \quad 1 > L(t, x_0) = \alpha > 0.$$

Denote, then, by  $\Omega$  the set of all points  $x$  with

$$(69) \quad L(t, x) \geq \alpha.$$

Obviously  $\Omega = \Omega_+$ , since  $p(x) = 0$  implies  $L(t, x) = 0$ . Now by (65) we have

$$L(t, x) \leq \int_{E_+} L(t, y) \Pi(x, dE_y),$$

and consequently

$$\begin{aligned} \alpha = L(t, x_0) &\leq \int_{E_+} L(t, y) \Pi(x_0, dE_y) \leq \alpha \Pi(x_0, E - \Omega) + \int_{\Omega} L(t, y) \Pi(x_0, dE_y), \\ &= \alpha \Pi_{\Omega}^{(1)}(x_0, E - \Omega) + \int_{\Omega} L(t, y) \Pi_{\Omega}^{(1)}(x_0, dE_y) \end{aligned}$$



so that by an inductive argument, using (56) we have

$$(70) \quad \alpha \leq \alpha \sum_{k=0}^n \Pi_{\Omega}^{(k)}(x_0, E - \Omega) + \int_{\Omega} L(t, y) \Pi_{\Omega}^{(n)}(x_0, dE_y).$$

Here the sign of equality can hold only if  $\Pi_{\Omega}^{(k)}(x_0, E - \Omega) = 0$  for  $k = 0, \dots, n$ . Since  $L(t, y) \leq 1$  we get from (70) using (57)

$$(71) \quad \alpha \leq \alpha(1 - \eta) + \eta,$$

where  $\eta$  is defined by

$$\Pi_{\Omega}^{(n)}(x_0, \Omega) \downarrow \eta.$$

Again, in (71) the sign of equality can hold only if  $\Pi_{\Omega}^{(k)}(x_0, E - \Omega) = 0$  for all  $k$ ; but by (57) we have in this case  $\eta = 1$ . Otherwise  $\alpha < \alpha(1 - \eta) + \eta$  so that certainly  $\eta > 0$ . Thus  $\Omega = \Omega_+$  is a set with the property stated in the theorem. However, by (67), (68) and (69) we get

$$\begin{aligned} \alpha \sum_{k=0}^{\infty} \int_{\Omega} \frac{1}{p(y)} \Pi^{(k)}(x_0, dE_y) &\leq \sum_{k=0}^{\infty} \int_{\Omega} \frac{1}{p(y)} L(t, y) \Pi^{(k)}(x_0, dE_y) \\ &\leq \sum_{k=0}^{\infty} \int_{E_+} \frac{1}{p(y)} L(t, y) \Pi^{(k)}(x_0, dE_y) \leq t, \end{aligned}$$

which means that the series (61) is convergent. Thus the divergence of the series implies  $\alpha = 0$  or  $P(t, x_0, E) = 1$ .

(ii) *Necessity.* This part of the proof will mainly use the representation of  $P(t, x, \Lambda)$  given in Theorem 1.

Suppose that condition (i) of the theorem does not hold, that is to say, that there is a set  $\Omega = \Omega_+$  for which (59) holds for some fixed  $x_0 \in \Omega$  and for which

$$(72) \quad \sum_{k=0}^{\infty} \int_{\Omega} \frac{1}{p(y)} \Pi_{\Omega}^{(k)}(x_0, dE_y) = a < \infty.$$

(72) implies in particular that all the integrals occurring converge. For this fixed set  $\Omega$  and all points  $x \in \Omega$  we define, in analogy to (27), an additive function  $P_{\Omega}(t, x, \Lambda)$  of sets  $\Lambda \in \mathfrak{F}$  by the recurrence formula

$$(73) \quad \begin{aligned} P_{\Omega}^{(0)}(t, x, \Lambda) &= P^{(0)}(t, x, \Lambda), \\ P_{\Omega}^{(n)}(t, x, \Lambda) &= \int_0^t d\sigma \int_E p(y) \Pi^*(t - \sigma, y, \Lambda \Omega) P_{\Omega}^{(n-1)}(\sigma, x, dE_y), \end{aligned}$$

where

$$\Pi^*(t, x, \Lambda) = \int_{\Lambda} \exp \{-tp(y)\} \Pi(x, dE_y).$$

Obviously

$$(74) \quad 0 \leq P_{\Omega}^{(n)}(t, x, \Lambda) \leq P^{(n)}(t, x, \Lambda),$$

and putting

$$(75) \quad W_{\Omega}^{(n)}(t, x, \Lambda) = P^{(n)}(t, x, \Lambda) - P_{\Omega}^{(n)}(t, x, \Lambda)$$

it is readily seen that both  $W_{\Omega}^{(n)}(t, x, \Lambda)$  and  $P_{\Omega}^{(n)}(t, x, \Lambda)$  are non-negative completely additive functions of sets  $\Lambda \in \mathfrak{B}$ . Furthermore

$$(76) \quad P(t, x, \Lambda) = \sum_{n=0}^{\infty} P_{\Omega}^{(n)}(t, x, \Lambda) + \sum_{n=0}^{\infty} W_{\Omega}^{(n)}(t, x, \Lambda).$$

$P^{(n)}(t, x, \Lambda)$  can be interpreted as the compound probability that the state  $X(t)$ , starting the point  $x \in \Omega$  at  $t=0$ , will during the time  $t$  change by exactly  $n$  jumps and in such a way that it remains contained in  $\Omega$  during the whole time and is contained in  $\Lambda$  at the moment  $t$ . Of course  $P_{\Omega}^{(n)}(t, x, E - \Omega) = 0$  for  $x \in \Omega$ .

Now (33) reads in our present notation

$$(77) \quad \begin{aligned} P^{(n+1)}(t, x, \Lambda) + \int_0^t d\sigma \int_{\Lambda} p(y) P^{(n+1)}(\sigma, x, dE_y) \\ = \int_0^t d\sigma \int_E p(y) \Pi(y, \Lambda) P^{(n)}(\sigma, x, dE_y). \end{aligned}$$

It is easily seen that the same calculations lead, for  $P_{\Omega}^{(n+1)}(t, x, \Lambda)$ , to the analogous formula (supposing, of course,  $x \in \Omega$ )

$$(78) \quad \begin{aligned} P_{\Omega}^{(n+1)}(t, x, \Lambda) + \int_0^t d\sigma \int_{\Lambda} p(y) P_{\Omega}^{(n+1)}(\sigma, x, dE_y) \\ = \int_0^t d\sigma \int_E p(y) \Pi(y, \Lambda \cap \Omega) P_{\Omega}^{(n)}(\sigma, x, dE_y). \end{aligned}$$

Subtracting (78) from (77), we get by (75)

$$(79) \quad \begin{aligned} W_{\Omega}^{(n+1)}(t, x, \Lambda) + \int_0^t d\sigma \int_{\Lambda} p(y) W_{\Omega}^{(n+1)}(\sigma, x, dE_y) \\ = \int_0^t d\sigma \int_E p(y) \Pi(y, \Lambda - \Lambda \cap \Omega) P_{\Omega}^{(n)}(\sigma, x, dE_y) \\ + \int_0^t d\sigma \int_E p(y) \Pi(y, \Lambda) W_{\Omega}^{(n)}(\sigma, x, dE_y). \end{aligned}$$

Now for any  $x \in \Omega$  and any  $\Lambda$

$$\int_0^t P_0^{(0)}(\sigma, x, \Lambda) d\sigma \leq \frac{1}{p(x)} \delta(x, \Lambda\Omega) = \int_{\Lambda+} \frac{1}{p(y)} \Pi_0^{(0)}(x, dE_y),$$

and from (78) we get

$$\int_{\Lambda} p(y) \int_0^t P_0^{(n+1)}(\sigma, x, dE_y) d\sigma \leq \int_E p(y) \Pi(y, \Lambda\Omega) \int_0^t P_0^{(n)}(\sigma, x, dE_y) d\sigma,$$

and thus by induction

$$(80) \quad \int_{\Lambda} p(y) \int_0^t P_0^{(n)}(\sigma, x, dE_y) d\sigma \leq \Pi_0^{(n)}(x, \Lambda\Omega)$$

or

$$(81) \quad \int_0^t P_0^{(n)}(\sigma, x, \Lambda) d\sigma \leq \int_{\Lambda+} \frac{1}{p(y)} \Pi_0^{(n)}(x, dE_y)$$

(the convergence of the right-hand member being guaranteed by (72)).

It follows from (81) and (72) that

$$\sum_{n=0}^{\infty} \int_0^t P_0^{(n)}(\sigma, x_0, \Omega) d\sigma \leq a,$$

and since  $P^{(n)}(\sigma, x_0, E - \Omega) = 0$  we can also write

$$(82) \quad \sum_{n=0}^{\infty} \int_0^t P_0^{(n)}(\sigma, x_0, E) d\sigma \leq a.$$

Next, we deduce a limitation for  $W_0^{(n)}(t, x_0, E)$ . Putting for  $n \geq 1$

$$\alpha_n = \int_0^t d\sigma \int_E p(y) \Pi(y, E - \Omega) P_0^{(n-1)}(\sigma, x_0, dE_y)$$

we readily get by (80)

$$0 \leq \alpha_n \leq \int_0^t \Pi(y, E - \Omega) \Pi_0^{(n-1)}(x_0, dE_y) = \Pi_0^{(n)}(x_0, E - \Omega).$$

Hence, by (57) and (59),

$$\sum_{n=1}^{\infty} \alpha_n \leq 1 - \alpha < 1.$$

By definition  $W_0^{(0)}(t, x_0, E) = 0$ . Hence we obtain from (79)

$$\sum_{n=0}^N W_0^{(n)}(t, x_0, E) = \sum_{n=1}^N \alpha_n - \int_0^t d\sigma \int_{E-\Omega} p(y) W_0^{(N)}(\sigma, x_0, dE_y) \leq 1 - \alpha$$

or

$$(83) \quad \sum_{n=0}^{\infty} W_{\Omega}^{(n)}(t, x_0, E) \leq 1 - \alpha.$$

Combining (83) with (82) we get finally, using (76),

$$\int_0^t P(\sigma, x_0, E) d\sigma \leq \alpha + (1 - \alpha)t$$

or, for  $t$  sufficiently large,

$$\int_0^t P(\sigma, x_0, E) d\sigma < t.$$

It follows that  $P(t, x_0, E) \neq 1$  which proves the necessity of our condition.

A few words may be added about the meaning of the conditions of Theorem 7. Suppose that there is a set  $\Omega = \Omega_+$  such that  $\Pi_{\Omega}^{(n)}(x_0, \Omega) > \eta > 0$  for some  $x_0 \in \Omega$ , and such that (61) converges. Consider, then, a random point moving at given moments in  $E$  by jumps according to the probability laws expressed by the ordinary Markoff chain (55). It is obvious that, if the point remained in  $\Omega$  during the first  $n$  steps, the probability of never leaving  $\Omega$  tends to 1. Thus, for any  $\epsilon > 0$ , there are points  $x \in \Omega$  for which  $\Pi_{\Omega}^{(n)}(x, \Omega) > 1 - \epsilon$ . The proof given for the sufficiency of our condition shows that also for all these points the series (61) will converge. Thus, in the statement of the Theorem 7  $\alpha$  can be replaced by  $1 - \epsilon$ .

Denote now, as before, by  $\Lambda_a$  the set of points  $x$ , with  $p(x) \leq a$ . The convergence of (61) implies the convergence of

$$\sum_{n=0}^{\infty} \Pi^{(n)}(x, \Lambda_a \Omega)$$

for any fixed  $a$ , and accordingly there is some sequence  $a_n \uparrow \infty$  such that even

$$\sum_{n=0}^{\infty} \Pi^{(n)}(x, \Lambda_{a_n} \Omega)$$

converges. This means, however, that there is a probability  $\eta > 0$  for our moving point to be contained for all  $n$  after  $n$  steps in  $\Omega - \Omega_{\Lambda_{a_n}}$ , that is to say, in a part of  $\Omega$ , where  $p(x) > a_n \uparrow \infty$ . In other words, there is a positive probability that our moving point will move, in the mean, towards points with increasing  $p(x)$ ; and if it did so for the first  $n$  steps, the probability that it will continue tends to unity as  $n \rightarrow \infty$ . Thus, in terms of the ergodic properties of the Markoff chain (55), the convergence of (61) is only possible if the point  $x$  is contained in a dissipative part of  $E$ .

Now the same reasoning applies also to the change of the state  $X(t)$  under

the influence of our stochastic process (cf. the interpretation of (55) on page 506). Roughly speaking, if  $P(t, x, E) < 1$ , the difference  $1 - P(t, x, E)$  can be interpreted as the probability that the state  $X$  will, starting from the point  $x$ , change during the time  $t$  by infinitely many jumps. The  $n$ th jump takes  $X$  in the set  $\Omega_n \subset \Omega$  and  $\Omega_n \rightarrow 0$ . It follows from Theorem 7 in particular that we have  $P(t, x, E) = 1$  for any point  $x$  belonging to an ergodic part of  $E$ —that is to say, if there is some bounded set  $\Delta$  such that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \Pi^{(n)}(x, \Delta) > 0.$$

### 7. Examples.

(i) Consider the case of an enumerable  $E$ , with the points  $x_0, x_1, \dots$ , and of a temporally homogeneous process. Let the  $p_i$  be any given positive constants and

$$\Pi_{ik} = \begin{cases} 1 & \text{for } k = i + 1, \\ 0 & \text{for } k \neq i + 1. \end{cases}$$

That is to say, from  $x_i$  only a transition to  $x_{i+1}$  is possible, and the probability of such a transition during an interval of length  $\Delta t$  is  $p_i \Delta t + o(\Delta t)$ .

The differential equations (13) of the process take on the form

$$(84) \quad P'_{ik}(t) = -p_k P_{ik}(t) + p_{k-1} P_{i, k-1}(t)$$

so that

$$(85) \quad P_{ik}(t) = 0 \text{ for } k < i, \quad P_{ii}(t) = e^{-p_i t}$$

and

$$P_{ik}(t) = p_{k-1} \int_0^t \exp\{-p_k^{(t-\sigma)}\} P_{i, k-1}(\sigma) d\sigma \text{ for } k > i.$$

In the case that  $p_i \neq p_k$  for  $i \neq k$ , the explicit solution is for  $k > i$

$$P_{ik}(t) = (-1)^{k-i} p_i p_{i+1} \cdots p_{k-1} \sum_{r=i}^k \frac{e^{-p_r t}}{(p_r - p_i)(p_r - p_{i+1}) \cdots (p_r - p_{r-1})(p_r - p_{r+1}) \cdots (p_r - p_n)};$$

it can be verified by means of the Lagrange interpolation formula but is of little use. The solution in the case in which  $p_i \neq p_k$  for  $i \neq k$  is not necessarily true, follows by the usual passage to the limit. We have

$$\Pi_{ik}^{(n)} = \begin{cases} 0 & \text{for } k \neq i + n \\ 1 & \text{for } k = i + n \end{cases}$$

and thus, by Theorem 7, the necessary and sufficient condition for  $\sum_k P_{ik}(t) = 1$  is that  $\sum_{n=0}^{\infty} 1/p_{i+n}$  diverges<sup>(20)</sup>. This can also be easily verified directly. Putting

$$(86) \quad L_{i,k}(t) = \int_0^t p_k P_{ik}(\sigma) d\sigma,$$

it follows from (84) for  $k > i$  that

$$(87) \quad P_{ik}(t) + L_{ik}(t) = L_{i,k-1}(t)$$

while  $P_{ii}(t) + L_{ii}(t) = 1$  and thus  $L_{i,i+n}(t) \downarrow L_i(t)$  as  $n \rightarrow \infty$ . On the other hand we have by (87)

$$\sum_{k=i}^{i+n} P_{ik}(t) = 1 - L_{i,i+n}(t)$$

or

$$(88) \quad \sum_{k=0}^{\infty} P_{ik}(t) = 1 - L_i(t) \leq 1.$$

But by (86) and (88),  $L_i(t) > 0$  would imply

$$(89) \quad t \geq \sum_{k=0}^{\infty} \int_0^t P_{ik}(\sigma) d\sigma \geq L_i(t) \sum_{k=i}^{\infty} \frac{1}{p_k},$$

that is to say, the convergence of  $\sum_k 1/p_{k+i}$ . Conversely, by (89) the divergence of  $\sum_k 1/p_{k+i}$  implies that  $L_i(t) = 0$ , or by (88) that  $\sum_k P_{ik}(t) = 1$ . A similar argument can be applied even in the case that the  $p_i$  depend on  $t$ .

The stochastic process just described plays an important role for different applications. In the case that all  $p_i$  are equal,  $p_i = p$ , it reduces to the classical Poisson process

$$P_{ik}(t) = e^{-pt} \frac{(pt)^{k-i}}{(k-i)!} \quad (k \geq i).$$

The general case was used by Lundberg [7] in the theory of invalidity insurance, and by Feller [4] to describe the growth of some biological populations. In both cases it is natural to assume that  $p_i \rightarrow \infty$  as  $i \rightarrow \infty$ . The same stochastic process was also applied to describe radioactive processes,  $x_i$  standing for the "elementary probability" of its disintegration; but here, of course, the space  $E$  contains only a finite number of points (or, what amounts to the same, some  $p_k = 0$ ).

(ii) Finally we give an example which proves that the condition of the corollary to Theorem 7 is not sufficient.

Let  $E$  consist of the points  $x_i$ ,  $i = 0, \pm 1, \pm 2, \pm 3, \dots$ . The process is

<sup>(20)</sup> The vanishing of any particular  $p_n$  obviously implies that  $P_{ik}(t) = 0$  for any couple  $(i, k)$  with  $i \leq n < k$ ; and it is readily seen that  $\sum_k P_{ik}(t) = 1$  for any  $i \leq n$ .



again temporally homogeneous. For  $i < 0$  only the transition  $x_i \rightarrow x_{i-1}$  is possible; for  $i \geq 0$  both  $x_i \rightarrow x_{i+1}$  and  $x_i \rightarrow x_{i-1}$  are possible, the corresponding probabilities being  $1 - \pi_i > 0$  and  $\pi_i > 0$ . In other words we suppose that

$$\Pi_{ik} = \begin{cases} 1 & \text{if } i < 0, k = i - 1, \\ 1 - \pi_i & \text{if } i \geq 0, k = i + 1, \\ \pi_i & \text{if } i \geq 0, k = -i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let us now suppose that (i) the product  $\prod (1 - \pi_i) = \alpha > 0$ , (ii)  $p_i = 1$  for  $i \leq 0$ , and (iii)  $\sum_{i=1}^{\infty} 1/p_i = a$  converges. Then the condition of the corollary to Theorem 7 is satisfied. For obviously we have if  $n > 0, i \geq 0$ ,

$$\Pi_{i, i-n}^{(n)} = \pi_i + (1 - \pi_i)\pi_{i+1} + \cdots + (1 - \pi_i)(1 - \pi_{i+1}) \cdots (1 - \pi_{i+n-1})\pi_{i+n}$$

and if  $i < 0$

$$\Pi_{i, i-n}^{(n)} = 1.$$

Hence

$$\sum_n \sum_k \frac{1}{p_k} \Pi_{ik}^{(n)} \geq \sum_n \Pi_{i, -|i|-n}^{(n)} \geq \sum_n \pi_i$$

diverges. But, taking for  $\Omega$  the set of all points  $x_i$  with  $i \geq 0$ , it is readily seen that (59) holds for any  $x = x_i, i \geq 0$ , and nevertheless the series converges.

#### REFERENCES

1. J. L. Doob, *Stochastic processes depending on a continuous parameter*, these Transactions, vol. 42 (1937), p. 107.
2. W. Dubrovski, *Eine Verallgemeinerung der Theorie der rein unstetigen stochastischen Prozesse von W. Feller*, Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS, vol. 19 (1938), p. 439.
3. W. Feller, *Zur Theorie der stochastischen Prozesse (Existenz- und Eindeigkeitssätze)*, Mathematische Annalen, vol. 113 (1936), p. 113.
4. ———, *Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung*, Acta Biotheoretica, vol. 5 (1939), p. 11.
5. M. Fréchet, *Recherches Théoriques Modernes sur le Calcul des Probabilités*, Part II (Traité du Calcul des Probabilités, vol. 1, no. 3), 1938.
6. A. Kolmogoroff, *Ueber die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Mathematische Annalen, vol. 104 (1931), p. 415.
7. O. Lundberg, forthcoming dissertation, Stockholm.
8. G. Pólya, *Sur la promenade au hasard dans un réseau des rues*, Lecture at the "Colloque Consacré à la Théorie des Probabilités," Geneva, 1937, Actualités Scientifiques et Industrielles, no. 734, 1938, p. 25.
9. Added in proof; cf. the footnote on page 492: W. Doeblin, *Sur certains mouvements aléatoires discontinus*, Skandinavisk Aktuarietidskrift, 1939, p. 211.

BROWN UNIVERSITY,  
PROVIDENCE, R. I.

## ON LINEAR TRANSFORMATIONS

BY

R. S. PHILLIPS

The purpose of this paper is to give a characterization of linear and completely continuous transformations both on the common Banach spaces to an arbitrary Banach space and vice versa. There is an abundant literature on this subject. Among the earliest papers, the now famous paper of Radon [24] should be mentioned. Here linear transformations on  $L^p$  to  $L^q$  ( $1 < p, q < \infty$ ) are characterized in a manner suggestive of the methods used in the present paper. The works of Gelfand [12], Dunford [6], Kantorovitch and Vulich [17], and Dunford and Pettis [9] contain much material on this subject supplementary to that treated here. In the interest of completeness we have restated a few of the results obtained by Gelfand [12], and Gowurin [13].

The principal tools used in our characterizations are certain abstractly valued function spaces. One such space is the class of all additive set functions  $x(\tau)$  on all Lebesgue measurable subsets  $\tau$  of  $(0, 1)$  to a Banach space  $X$  where for all linear functionals  $\bar{x}$  on  $X$  and for all subdivisions  $\pi = (\tau_1, \tau_2, \dots, \tau_n, \dots)$  of  $(0, 1)$  into disjoint sets,

$$\text{L.U.B.} \left[ \sum_{\tau} \frac{|\bar{x}[x(\tau_i)]|^q}{|\tau_i|^{q-1}} \right] < \infty.$$

If  $\phi(t) \in L^p$  ( $1/p + 1/q = 1$ ), we define an integral  $\int \phi dx$  to be the generalized  $\pi$ -limit of the unconditionally convergent sums  $\sum \phi(t_i)x(\tau_i)$  where  $t_i \in \tau_i$ . The function  $U(\phi) = \int \phi dx$  so defined on  $L^p$  is a characterization of the general linear transformation on  $L^p$  to  $X$ .

The first section is a study of the abstractly valued function spaces which will be used to characterize the transformations. Section 2 is devoted to a discussion of three different types of integrals needed in these characterizations. In §3 a necessary and sufficient condition for a subset of a Banach space  $Y$  to be conditionally compact is given in terms of an arbitrary determining manifold  $\Gamma$  in the conjugate space  $\bar{Y}$ . As a consequence, if a transformation  $U$  is additive and homogeneous on  $X$  to  $Y$  and its adjoint is completely continuous on  $\Gamma$  to  $\bar{Y}$ , then  $U$  is completely continuous on  $X$  to  $Y$ . The section also contains a characterization of conditionally compact sets in a Banach space by means of a generalized base. This is applied to the spaces  $L^p$  ( $1 \leq p \leq \infty$ ) in §§5 and 6. Section 4 contains the principal results of this paper, namely, a characterization for the classes of transformations considered. In §5 we

Presented to the Society, December 29, 1939; received by the editors March 9, 1940. This paper was received by the editors of the Annals of Mathematics November 18, 1939, accepted by them, and later transferred to these Transactions.

obtain representations by means of a kernel of the general completely continuous transformation and weakly completely continuous separable transformation on  $L$  to an arbitrary Banach space. By means of this result and a theorem due to Dunford and Pettis [9], we show that  $U^2$  is completely continuous whenever  $U$  on  $L$  to  $L$  is weakly completely continuous. As a further application of this work, we prove in §6 that completely continuous transformations on the spaces  $L^p$ ,  $l^p$ ,  $C$ ,  $c_0$ ,  $M_T$  ( $1 \leq p \leq \infty$ ) to an arbitrary Banach space are approximable in the norm by degenerate transformations. A final section is devoted to the extension of linear transformations.

We will consider an abstract class  $T$  of elements  $t$  possessing a sigma family  $\mathcal{G}$  of subsets  $\tau$  of  $T$ .  $\alpha(\tau)$  will be a single-valued, non-negative, completely additive measure function on  $\mathcal{G}$ , which need not be finite valued. It will be convenient to designate by  $|\tau|$  the value  $\alpha(\tau)$ .  $X$  will denote a Banach space of elements  $x$  and  $\bar{X}$  its conjugate space of elements  $\bar{x}$  [1, chap. 5]. We define with Dunford [7, p. 316] a determining manifold  $\Gamma$  in  $\bar{X}$  to be a closed linear subset of  $\bar{X}$  such that<sup>(1)</sup> L.U.B.  $[|\bar{x}(x)| \mid x \in X, \bar{x} \in \Gamma, \|\bar{x}\| \leq 1] = \|x\|$ .  $M_T$  will be the Banach space of bounded functions  $a(t)$  on an abstract class  $T = \{t\}$  to real numbers having the norm  $\|a\| = \text{L.U.B. } [|a(t)| \mid t \in T]$ .  $\pi$  will have three different meanings: type 1,  $\pi$  will be a finite or denumerable set of disjoint sets  $\tau$  of  $\mathcal{G}$  such that  $0 < |\tau| < \infty$ .  $\pi_1 \geq \pi_2$  will mean that  $\sum \tau^1 \supset \sum \tau^2$ , and that every set  $\tau^1 \in \pi_1$  is either a subset of some  $\tau^2 \in \pi_2$  or  $\tau^1$  is disjoint from every  $\tau^2 \in \pi_2$ ; type 2,  $\pi$  will be a subdivision of  $T$  into a finite number of disjoint sets  $\tau \in \mathcal{G}$  ( $\mathcal{G}$  need not possess a measure function).  $\pi_1 \geq \pi_2$  will mean that each  $\tau^1 \in \pi_1$  is a subset of some  $\tau^2 \in \pi_2$ ; and type 3,  $\pi$  will be a subdivision of the interval  $(0, 1)$  into a finite number of intervals the maximum of whose lengths is  $|\pi|$ .  $\pi_1 \geq \pi_2$  will mean that  $|\pi_1| \leq |\pi_2|$ . In each case the relation  $\geq$  on the class  $[\pi]$  is transitive and compositive. The general limit of E. H. Moore-H. L. Smith [20, p. 103] can therefore be defined on each of these ranges.  $\text{Lim}_\pi$  will designate this limit.

1. **Abstractly valued function spaces.** We will be interested in the following classes of functions:

$$\begin{aligned} V^1(X, \Gamma) &= \left[ x(\tau) \mid \text{L.U.B. } \sum_{\tau} |\bar{x}[x(\tau_i)]| < \infty, \bar{x} \in \Gamma \right], \\ V^q(X, \Gamma) &= \left[ x(\tau) \mid \text{L.U.B. } \sum_{\tau} \frac{|\bar{x}[x(\tau_i)]|^q}{|\tau_i|^{q-1}} < \infty, \bar{x} \in \Gamma \right], & 1 < q < \infty, \\ V^\infty(X) &= [x(\tau) \mid \|x(\tau)\| \leq M \cdot |\tau|, \tau \in \mathcal{G}, |\tau| < \infty], \\ v^q(X, \Gamma) &= \left[ x_n \mid \sum_n |\bar{x}(x_n)|^q < \infty, \bar{x} \in \Gamma \right], & 1 \leq q < \infty, \\ v^\infty(X) &= \left[ x_n \mid \text{L.U.B. } \|x_n\| < \infty \right]. \end{aligned}$$

(<sup>1</sup>) The class of elements  $s$  satisfying the property  $P$  will be designated by  $[s|P]$ .

For  $q \neq 1$  ( $q=1$ )  $\pi$  is always to be understood as being of type 1 (type 2). For  $q=1$ ,  $\mathfrak{G}$  may be a finitely additive Jordan field. It will be convenient to denote an element of one of these function classes by  $\hat{x}$ .

If  $T$  is the set of positive integers,  $\mathfrak{G}$  the family of all subsets of  $T$ , and if  $|\tau|$  is equal to the number of integers in  $\tau$ , then  $V^q(X, \Gamma)$  ( $1 < q < \infty$ ) and  $V^\infty(X)$  are identical with  $v^q(X, \Gamma)$  and  $v^\infty(X)$  respectively. Theorems analogous to Theorems 1.1, 1.2, and 1.5 have been proved for  $v^1(X, \Gamma)$  by Gelfand [12].

1.1. THEOREM. If  $\hat{x} \in V^q(X, \Gamma)$ ,  $1 \leq q < \infty$ , then there exists an  $M$  such that<sup>(2)</sup>

$$\text{L.U.B.} \left[ \sum_{\tau} \frac{|\hat{x}[x(\tau_i)]|^q}{|\tau_i|^{q-1}} \right]^{1/q} \leq M \cdot \|\hat{x}\|$$

for all  $\hat{x} \in \Gamma$ .

Define

$$p(\hat{x}) = \text{L.U.B.} \left[ \sum_{\tau} \frac{|\hat{x}[x(\tau_i)]|^q}{|\tau_i|^{q-1}} \right]^{1/q}$$

on  $\Gamma$ . It is easy to show that  $p(\hat{x}) \geq 0$ ,  $p(\hat{x}_1 + \hat{x}_2) \leq p(\hat{x}_1) + p(\hat{x}_2)$ , and that  $\hat{x}_n \rightarrow \hat{x}$  implies  $\liminf_{n \rightarrow \infty} p(\hat{x}_n) \geq p(\hat{x})$ . The theorem now follows from a lemma due to Gelfand [12, p. 240].

We define a norm for the several spaces as follows:

$$\|\hat{x}\| = \text{L.U.B.}_{\|\hat{x}\|=1} \left\{ \text{L.U.B.}_{\tau} \left[ \sum_{\tau} \frac{|\hat{x}[x(\tau_i)]|^q}{|\tau_i|^{q-1}} \right]^{1/q} \mid \hat{x} \in \Gamma \right\},$$

$$\hat{x} \in V^q(X, \Gamma), \quad 1 \leq q < \infty,$$

$$\|\hat{x}\| = \text{L.U.B.} \left[ \frac{\|\hat{x}(\tau)\|}{|\tau|} \mid \tau \in \mathfrak{G}, |\tau| < \infty \right],$$

$$\hat{x} \in V^\infty(X),$$

$$\|\hat{x}\| = \text{L.U.B.} \left[ \sum_n |\hat{x}(x_n)| \mid \hat{x} \in \Gamma \right],$$

$$\hat{x} \in v^1(X, \Gamma).$$

1.2. THEOREM.  $V^q(X, \Gamma)$ ,  $1 \leq q \leq \infty$ , is a Banach space.

It is clear that the spaces are linear normed spaces. Only the proof of completeness remains. Suppose  $\{\hat{x}_n\}$  is a Cauchy sequence in  $V^q(X, \Gamma)$ ,  $1 \leq q < \infty$ . Then for every  $\tau \in \mathfrak{G}$ ,

$$\lim_{m, n \rightarrow \infty} \frac{|\hat{x}_n(\tau) - \hat{x}_m(\tau)|^q}{|\tau|^{q-1}} = 0$$

uniformly in the unit sphere of  $\Gamma$ . Hence  $\lim_{m, n \rightarrow \infty} \|\hat{x}_n(\tau) - \hat{x}_m(\tau)\| = 0$  and there exists an additive set function  $x(\tau) = \lim_{n \rightarrow \infty} x_n(\tau)$ . Further, if we are given  $\hat{x}$  in

<sup>(2)</sup> For  $q=1$ , we define  $|\tau|^{q-1}$  to be identically one.

the unit sphere of  $\Gamma$ ,  $\pi$ , and  $N$ , then

$$\left[ \sum_1^N \frac{|\hat{x}[x(\tau_i)]|^q}{|\tau_i|^{q-1}} \right]^{1/q} \leq \lim_n \|\hat{x}_n\|$$

so that  $\hat{x} \in V^q(X, \Gamma)$ . Finally for an arbitrary  $\epsilon > 0$  there exists  $N_\epsilon$  such that if  $m, n \geq N_\epsilon$  then  $\|\hat{x}_m - \hat{x}_n\| \leq \epsilon$ . Therefore if  $n > N_\epsilon$

$$\|\hat{x} - \hat{x}_n\| = \text{L.U.B.}_{\|\tau\|=1, \tau \in N} \left\{ \lim_m \left[ \sum_1^N \frac{|\hat{x}[x_n(\tau_i) - x_m(\tau_i)]|^q}{|\tau_i|^{q-1}} \right]^{1/q} \right\} \leq \epsilon.$$

Completeness for  $V^\infty(X)$  can be demonstrated in a similar fashion.

If  $X$  is the space of real numbers  $R$ , then  $\Gamma$  must be identical with  $\bar{X} = R$ . For  $T = \sum \tau_i$  ( $|\tau_i| < \infty$ ) it is well known that  $V^q(R)$  ( $1 \leq q \leq \infty$ ) is equivalent [1, p. 180] to the space  $L^q(\alpha)$  of measurable functions  $\psi(t)$  for which  $\int_T |\psi(t)|^q d\alpha < \infty$  ( $1 \leq q < \infty$ ) and  $\text{ess. L.U.B. } [|\psi(t)| | t \in T] < \infty$  ( $q = \infty$ ). The equivalence is defined by the transformation  $U(\psi) = x(\tau) = \int \psi(t) d\alpha$  for  $|\tau| < \infty$ .

A sum  $\sum x_n$  will be said to be unconditionally convergent if  $\sum x_n$  summed over any subsequence of the integers converges.  $x(\tau)$  will be called completely additive if for any sequence of disjoint measurable sets  $\{\tau_n\}$ ,  $\sum x(\tau_n) = x(\sum \tau_n)$  where the sum is unconditionally convergent.

**1.3. THEOREM.** If  $\hat{x} \in V^q(X, \Gamma)$  ( $1 < q \leq \infty$ ) and if  $\tau_0$  is of finite measure, then  $x(\tau)$  is an absolutely continuous and completely additive set function on measurable subsets of  $\tau_0$ .

If  $\hat{x} \in V^q(X, \Gamma)$ , then

$$\left[ \frac{|\hat{x}[x(\tau)]|^q}{|\tau|^{q-1}} \right]^{1/q} \leq \|\hat{x}\| \cdot \|\tau\|$$

for all  $\hat{x} \in \Gamma$  and all  $\tau$  ( $0 < |\tau| < \infty$ ). Hence  $\|x(\tau)\| \leq \|\hat{x}\| \cdot |\tau|^{1-1/q}$  which implies absolute continuity. Let us now consider  $x(\tau)$  and  $V^q(R)$  on subsets of a set  $\tau_0$  of finite measure. Since  $\hat{x}[x(\tau)] \in V^q(R)$  it follows from the above that there exists  $\psi(t) \in L^q(\alpha)$  for which  $\hat{x}[x(\tau)] = \int \psi(t) d\alpha$ . Given a sequence  $\{\tau_n\}$  of disjoint measurable sets, then

$$\hat{x}[x(\sum \tau_n)] = \int_{\sum \tau_n} \psi d\alpha = \sum \int_{\tau_n} \psi d\alpha = \sum \hat{x}[x(\tau_n)].$$

By a theorem due to Dunford [7, p. 326, Theorem 32],  $x(\sum \tau_n) = \sum x(\tau_n)$  which is unconditionally convergent.

It is clear that the transformation  $U(\hat{x}) = \hat{x}[x(\tau)]$  on  $\Gamma$  to  $V^q(R)$  ( $1 \leq q \leq \infty$ ) is linear and that  $\|U\| = \|\hat{x}\|$ . We define  $V^q_c(X, \Gamma)$  to be the subspace of  $V^q(X, \Gamma)$  for which this transformation is completely continuous. Because the class of completely continuous transformations on  $X$  to  $Y$  is a

closed linear subspace of the space of linear transformations, it follows that the same is true of  $V^q_c(X, \Gamma)$  in  $V^q(X, \Gamma)$ . We define  $v^q_c(X, \Gamma)$  ( $1 \leq q \leq \infty$ ) in a similar fashion.

1.4. THEOREM. *A necessary and sufficient condition that  $\hat{x}$  belong to  $V^\infty_c(X)$  is that  $\hat{x}$  belong to  $V^\infty(X)$  and that the set  $[x(\tau)/|\tau| \mid \tau \in \mathcal{G}]$  be conditionally compact.*

This is an immediate consequence of Theorem 3.1.

1.5. THEOREM. *A necessary and sufficient condition that  $\hat{x}$  belong to  $V^1_c(X, \Gamma)$  is that  $\hat{x}$  belong to  $V^1(X, \Gamma)$  and that the set  $[x(\tau) \mid \tau \in \mathcal{G}]$  be conditionally compact.*

This again follows from Theorem 3.1.

1.6. THEOREM. *If  $T = (0, 1)$  and  $\alpha$  is the Lebesgue measure function, then for  $1 < q < \infty$  the following are equivalent statements:*

- (1)  $\hat{x} \in V^q_c(X, \Gamma)$ .
- (2)  $\hat{x} \in V^q(X, \Gamma)$  and

$$\lim_{h \rightarrow 0} \int_0^1 \left| \frac{d\hat{x}[x(I_0^t)]}{dt} \Big|_{t+h} - \frac{d\hat{x}[x(I_0^t)]}{dt} \Big|_t \right|^q dt = 0$$

uniformly for all  $\hat{x}$  in the unit sphere of  $\Gamma$  ( $I_0^t = (0, t)$ ).

This follows from the above remarks on the equivalence between  $V^q(R)$  and  $L^q$  and a theorem due to M. Riesz on compact sets in  $L^q$  [25].

1.7. THEOREM. *A necessary and sufficient condition that  $\hat{x}$  belong to  $v^q_c(X, \Gamma)$  for  $1 \leq q < \infty$  is that  $\hat{x}$  belong to  $v^q(X, \Gamma)$  and that  $\lim_{n \rightarrow \infty} \sum_n |\hat{x}(x_i)|^q = 0$  uniformly in the unit sphere of  $\Gamma$ .*

This follows from a well known theorem on compact sets in  $l^q$ , which for  $q=2$  is due to Fréchet [11, p. 19].

If  $\hat{x} \in v^1(X, \Gamma)$ , then  $\|\sum_n x_n\| \leq \|\hat{x}\| \leq 2 \cdot \text{L.U.B.}_\nu \|\sum_n x_n\|$  where  $\nu$  runs through all finite sets of integers. If  $\hat{x} \in v^1_c(X, \Gamma)$ , we have as a corollary to Theorem 1.7 that  $\sum_n x_n$  is unconditionally convergent. Dunford has shown that if  $\sum_n x_n$  is unconditionally convergent, then  $\hat{x} \in v^1_c(X, \Gamma)$  [7, p. 326, Theorem 32].

1.8. THEOREM. *If  $\bar{X}$  is separable and if  $\hat{x} \in v^1(\bar{X}, \Gamma)$ , then  $\hat{x} \in v^1_c(\bar{X}, \bar{X})$ .*

By hypothesis  $\sum |\hat{x}_n(x)| \leq \|\hat{x}\| \cdot \|x\|$  for every  $x \in \bar{X}$ . Hence for every denumerable set of integers  $\sigma$ ,  $|\sum_\sigma \hat{x}_n(x)| \leq \|\hat{x}\| \cdot \|x\|$  and is therefore a linear functional  $\hat{x}_\sigma$  on  $X$ . Since  $X$  is a determining manifold in  $\bar{X}$ , it follows by the above mentioned Dunford theorem that  $\sum x_n$  is unconditionally convergent. In other words  $\hat{x} \in v^1_c(\bar{X}, \bar{X})$ .



1.9. COROLLARY. If  $X$  is separable and there exists an  $\hat{x} \in v^1(X, \bar{X})$  which is not an element of  $v^1_0(X, \bar{X})$ , then  $X$  and any separable space containing  $X$  as a subspace are not conjugate spaces.

This permits another demonstration of the fact that  $c$  and hence  $C$  which contains  $c$  is not a conjugate space. Let  $x_n$  be the  $n$ th unit vector in  $c$ . Any  $\hat{x} = \{a_n\} \in l$  [1, p. 67].  $\sum |\hat{x}(x_n)| = \sum |a_n| < \infty$  implies that  $\hat{x} \in v^1(c, l)$ . However  $\sum x_n$  is obviously not unconditionally convergent.

2. Integrals. It is convenient to divide the discussion of this section into three parts: (1) an integral involving functions  $\phi(t) \in L^p(\alpha)$  and  $\hat{x} \in V^q(X, \Gamma)$  where  $1/p + 1/q = 1$  and  $1 < q \leq \infty$ ; (2) an integral involving functions  $\phi(t)$  either bounded or  $\alpha$ -measurable and essentially bounded and  $\hat{x} \in V^1(X, \Gamma)$ ; and (3) an integral involving functions  $\phi(t) \in C$  and functions  $x(t)$  to be defined.

2.1. DEFINITION. For functions  $\phi(t) \in L^p(\alpha)$  and  $\hat{x} \in V^q(X, \Gamma)$ ,  $1 < q \leq \infty$ , we define

$$\int \phi dx = \lim_{\pi} \sum_{\tau_i} \phi(t_i) x(\tau_i) \quad (\pi \text{ of type 1})$$

whenever for some  $\pi_0$  and all  $\pi \geq \pi_0$ ,  $\sum_{\tau_i} \phi(t_i) x(\tau_i)$  is unconditionally convergent for each  $t_i \in \tau_i$  and the limit exists.

The multiple valued function  $x(\pi)$  on a transitive and compositive class  $[\pi]$  will be said to be a fundamental  $\pi$ -sequence if for an arbitrary  $\epsilon > 0$  there exists a  $\pi_\epsilon$  such that for  $\pi_1, \pi_2 \geq \pi_\epsilon$ ,  $\|x(\pi_1) - x(\pi_2)\| \leq \epsilon^{(*)}$ .

2.2. LEMMA. If  $[x(\pi)]$  is a fundamental  $\pi$ -sequence and  $U$  a linear transformation on  $X$  to  $Y$ , then there exists an  $x \in X$  such that  $x = \lim_{\pi} x(\pi)$  and  $U(x) = \lim_{\pi} U[x(\pi)]^{(*)}$ .

Choose a sequence of positive numbers  $\epsilon_n \rightarrow 0$ . It is clear that one can obtain a sequence  $\pi_n$  such that  $\pi_{n+1} \geq \pi_n$  and for  $\pi \geq \pi_n$ ,  $\|x(\pi) - x(\pi_n)\| \leq \epsilon_n$ . Let  $x'(\pi_n)$  be one of the elements of  $x(\pi_n)$ . As  $X$  is sequentially complete there will exist an  $x \in X$  such that  $x = \lim_n x'(\pi_n)$ . But then if  $\pi \geq \pi_n$ ,  $\|x - x(\pi)\| \leq 2\epsilon_n$  and likewise  $\|U(x) - U[x(\pi)]\| \leq \|U\| \cdot 2\epsilon_n$ . Hence  $x = \lim_{\pi} x(\pi)$  and  $U(x) = \lim_{\pi} U[x(\pi)]$ .

2.3. THEOREM. If  $\phi(t) \in L^p(\alpha)$  and  $\hat{x} \in V^q(X, \Gamma)$ , then  $\int \phi dx$  exists.

Given  $\epsilon > 0$ , there exists  $\pi_\epsilon$  such that if  $\pi \geq \pi_\epsilon$  then  $\sum_{\tau_i} |\phi(t_i)|^p |\tau_i| < \infty$  and  $|\sum_{\tau_i} [\phi(t_i) - \phi(t'_i)]^p |\tau_i|| \leq \epsilon$  for all  $t_i \in \tau_i \in \pi$  and all  $t'_i \in \tau'_i \in \pi$ , where  $\tau_i \subset \tau'_i$ . Then

(\*) If  $B$  is a subset of  $X$ ,  $\|B\| = \text{L.U.B. } [\|x\| \mid x \in B]$ .

(\*) Compare with Moore and Smith [20, p. 106].

$$\left| \sum_n \phi(t_i) \bar{x}[x(\tau_i)] \right| \leq \| \dot{x} \| \cdot \| \bar{x} \| \cdot \left[ \sum_n |\phi(t_i)|^p |\tau_i| \right]^{1/p}$$

and therefore approaches zero uniformly in the unit sphere of  $\Gamma$  as  $n \rightarrow \infty$ . It follows that  $\sum_n \phi(t_i) x(\tau_i)$  is unconditionally convergent. Further if  $\pi \geq \pi_*$ ,

$$\begin{aligned} \left\| \sum_{\tau} \phi(t_i) x(\tau_i) - \sum_{\tau_i} \phi(t_{ij}) x(\tau_{ij}) \right\| &= \text{L.U.B.}_{\substack{x \in \Gamma, \|x\|=1}} \left| \sum_{\tau} (\phi(t_i) - \phi(t'_{ij})) \bar{x}[x(\tau_i)] \right| \\ &\leq \| \dot{x} \| \cdot \left[ \sum_{\tau} |\phi(t_i) - \phi(t'_{ij})|^p |\tau_i| \right] \leq \| \dot{x} \| \cdot \epsilon. \end{aligned}$$

Hence  $\sum_n \phi(t_i) x(\tau_i)$  is a fundamental  $\pi$ -sequence and by Lemma 2.2  $\lim_n \sum_n \phi(t_i) x(\tau_i) = \int \phi dx$  exists.

2.4. THEOREM. L.U.B.  $[ \| \int \phi dx \| | \phi(t) \in L^p(\alpha), \| \phi \| = 1 ] = \| \dot{x} \|$ .

Since  $\bar{x}(\dot{x}) \in V^q(R)$  ( $\bar{x}(\dot{x})$  will be used to indicate the function  $\bar{x}[x(\tau)] \in V^q(R)$ )

$$\text{L.U.B.}_{\| \phi \| = 1} \left| \bar{x} \int \phi dx \right| = \text{L.U.B.}_{\| \phi \| = 1} \left| \int \phi d\bar{x}(x) \right| = \| \bar{x}(\dot{x}) \|.$$

Therefore

$$\begin{aligned} \text{L.U.B.}_{\| \phi \| = 1} \left\| \int \phi dx \right\| &= \text{L.U.B.} \left[ \left\| \bar{x} \int \phi dx \right\| \mid \bar{x} \in \Gamma, \| \bar{x} \| = 1, \| \phi \| = 1 \right] \\ &= \text{L.U.B.} [ \| \bar{x}(\dot{x}) \| \mid \bar{x} \in \Gamma, \| \bar{x} \| = 1 ] = \| \dot{x} \|. \end{aligned}$$

One can likewise define this integral on any measurable set. We designate the so-defined integral on  $\tau$  by  $\int_{\tau} \phi dx$ .  $\int_{\tau} \phi dx$  is clearly an additive set function on  $\mathcal{G}$ . Since by Theorem 2.4  $\| \int_{\tau} \phi dx \| \leq \| \dot{x} \| \cdot [ \int_{\tau} |\phi|^p d\alpha ]^{1/p}$ , it follows that it is absolutely continuous and consequently completely additive.

2.5. THEOREM. If  $\phi(t) \in L^p(\alpha)$ ,  $\alpha(T) < \infty$ , and  $\bar{x} \in V^q(X, \Gamma)$  is such that  $x(\tau) = \int'_{\tau} y(t) d\alpha$ , then  $\int \phi dx = \int' \phi y d\alpha$  where  $\int'$  are both either Dunford integrals with  $y(t) \in \mathcal{L}_0^q(E)[X, \Gamma]$  [7] or Birkhoff integrals [3].

Suppose that  $y(t) \in \mathcal{L}_0^q(E)[X, \Gamma]$ ; then  $x(x(\tau)) = \int_{\tau} x(y) d\alpha$  for every  $\bar{x} \in \Gamma$ . It follows from the similar theorem in real variables that

$$\bar{x} \left[ \int_{\tau} \phi dx \right] = \int_{\tau} \phi d\bar{x}(x) = \int_{\tau} \phi \bar{x}(y) d\alpha.$$

This is equivalent to the statement that  $\phi(t) \cdot y(t) \in \mathcal{L}_0^1(E)[X, \Gamma]$  and  $\int \phi dx = \int' \phi y d\alpha$ .

On the other hand, suppose  $y(t)$  is Birkhoff integrable to the value  $x(\tau)$ . Then let  $\phi_n(t) = \phi(t)$  if  $|\phi(t)| \leq n$  and vanish elsewhere.  $\phi_n(t)y(t)$  differs from

$\phi(y)t(y)$  on a set whose measure approaches zero as  $n \rightarrow \infty$ . Since  $\phi_n(t)$  is bounded,  $\phi_n(t) \cdot y(t)$  is Birkhoff integrable [3, p. 369, Theorem 17]. Moreover it is integrable to the same value as the above Dunford integral so that  $\int \phi_n y d\alpha = \int \phi_n dx$ . By Theorem 2.4,

$$\left\| \int \phi_n y d\alpha \right\| = \left\| \int \phi_n dx \right\| \leq \|\dot{x}\| \cdot \left[ \int |\phi_n|^p d\alpha \right]^{1/p} \leq \|\dot{x}\| \cdot \left[ \int |\phi|^p d\alpha \right]^{1/p}.$$

The integrals  $\int \phi_n y d\alpha$  are therefore uniformly absolutely continuous. By a theorem due to the author [23, Theorem 6.2]  $\phi(t) \cdot y(t)$  is Birkhoff integrable and

$$\int' \phi y d\alpha = \lim_n \int \phi_n dx = \int \phi dx.$$

We will next consider  $\mathfrak{T}$ -measurable functions  $\phi(t)$  either bounded or essentially bounded relative to a measure  $\alpha$ .  $x(\tau) \in V^1(X, \Gamma)$  is defined on all  $\tau \in \mathfrak{T}$  and in the latter case vanishes on the null sets of  $\alpha$ . For convenience we will limit ourselves to the former case.

2.6. DEFINITION. For a bounded function  $\phi(t)$  and  $\dot{x} \in V^1(X, \Gamma)$ , we define

$$\int \phi dx = \lim_{\tau} \sum \phi(t_i) x(\tau_i) \quad (\pi \text{ of type 2})$$

whenever for  $t_i$  an arbitrary element of  $\tau$ , this limit exists.

When  $\int \phi dx$  exists by both Definition 2.1 and Definition 2.6, the value in each case is the same.

The following two theorems are special cases of a theorem due to Gowurin [13, pp. 265-266]. We omit their proofs.

2.7. THEOREM. If  $\phi(t)$  is bounded and  $\dot{x} \in V^1(X, \Gamma)$ , then  $\int \phi dx$  exists.

2.8. THEOREM. L.U.B.  $[\|\int \phi dx\| \mid |\phi(t)| \leq 1] = \|\dot{x}\|$ .

It is unlikely that much can be said about the differentiation of  $\dot{x} \in V^q(X, \overline{X})$  for  $1 \leq q < \infty$ . Pettis has constructed an  $\dot{x} \in V^2_c(L^2, L^2)$  [22, Example 9.4] which has no pseudo-derivative [22, p. 300]. In §5 we demonstrate that  $\dot{x} \in V^{\infty}_c(X)$  ( $X$  arbitrary) and  $\dot{x} \in V^{\infty}(X)$  ( $X$  separable and regular) for  $T = \sum \tau_i$  ( $|\tau_i| < \infty$ ) can be expressed as the Birkhoff integral of a function on  $T$  to  $X$ .

We wish finally to consider an integral for functions  $\phi(t) \in C$ . In this connection Gelfand [12, pp. 246-253] has introduced the abstractly valued function classes  $V(X)$  and  $V_c(X)$ .  $V(X)$  is the class of all functions  $x(t)$  on  $(0, 1)$  to  $X$  for which  $\dot{x}[x(t)]$  is of bounded variation and continuous on the left, while  $V_c(X)$  is the subclass of  $V(X)$  for which the set

$$[\sum (x(t_i) - x(t'_i)) \mid (t_i, t'_i) \text{ disjoint intervals}]$$

is compact. The L.U.B. [variation of  $\hat{x}[x(t)] \mid \|\hat{x}\| = 1$ ] exists and can be defined to be the norm  $\|\hat{x}\|$  for elements of  $V(X)$ . It is easily shown that  $V(X)$  is a Banach space having  $V_c(X)$  as a closed linear subspace.

2.9. DEFINITION. For functions  $\phi(t) \in C$  and  $\hat{x} \in V(X)$  we define

$$\int \phi d\hat{x} = \lim_{|\pi| \rightarrow 0} \sum_{\pi} \phi(t'_i) [x(t_i) - x(t_{i-1})] \quad (\pi \text{ of type 3})$$

whenever for  $t'_i$  an arbitrary element of  $(t_i, t_{i+1})$  the limit exists.

2.10. THEOREM. If  $\phi(t) \in C$  and  $\hat{x} \in V(X)$ , then  $\int \phi d\hat{x}$  exists and L.U.B.  $[\|\int \phi d\hat{x}\| \mid \|\phi(t)\| \leq 1] = \|\hat{x}\|$ .

This theorem has likewise been proven by Gowurin [13]. It is clear that  $\hat{x}[\int \phi d\hat{x}] = \int \phi d\hat{x}(x)$  so that this integral when it exists is equal to the integral defined by Gelfand [12, pp. 259-260].

3. On conditionally compact sets in a Banach space. In this section we will consider two different characterizations of conditionally compact sets in a Banach space  $X$ . The first is given in terms of a determining manifold  $\Gamma$ , while the second involves the notion of a generalized base.

3.1. THEOREM. A necessary and sufficient condition that the set  $S \equiv [x]$  be conditionally compact is that both L.U.B.  $[\|\hat{x}(x)\| \mid x \in S] < \infty$  for each  $\hat{x} \in \Gamma$  and  $U(\hat{x}) = \hat{x}(x)$  on  $\Gamma$  to  $M_S^{(b)}$  be completely continuous.

Let  $x_n$  be any denumerable subset of  $S$ . Its linear closed extension  $Y$  is a separable Banach space. Let  $\Gamma_1$  be the set of elements of  $\Gamma$  considered as members of the conjugate space of  $Y$ .  $\Gamma_1$  is clearly a determining manifold in the conjugate space of  $Y$ . The unit sphere of the conjugate space of a separable Banach space is a compact metric space in its weak topology [1, p. 186]. Hence  $\Gamma_1$  contains a denumerable subset  $\{\hat{x}_p\}$  which is weakly dense in  $\Gamma_1$ . The linear transformation  $V(x) = \hat{x}_p(x)$  on  $Y$  to  $m$  defines an equivalence. It is therefore sufficient to show that the set  $\{\hat{x}_p(x_n)\}$  is conditionally compact in  $m$ . By the diagonal procedure we can obtain an  $n$ -subsequence  $\hat{x}_p(x_{n'})$  such that  $\lim_{n'} \hat{x}_p(x_{n'})$  exists for every  $p$ . Moreover this limit exists uniformly in  $p$ . For if the contrary were true there would exist a  $p$ -subsequence having no subsequence for which the limit existed uniformly. As  $\|\hat{x}_p\| \leq 1$  and as  $U$  is completely continuous, this  $p$ -subsequence would have a subsequence  $p'$  for which  $\lim_{p'} \hat{x}_{p'}(x_{n'})$  exists uniformly in  $p'$  which gives a contradiction.

To prove the necessity we notice that the closed linear extension  $Y$  of  $S$  is a separable Banach space. Hence every bounded sequence of functionals

<sup>(b)</sup> We remind the reader that  $M_S$  is the space of bounded functions  $\hat{x}(x)$  on  $S$  to real numbers.

on  $Y$  contains a weakly convergent subsequence [1, p. 123, Theorem 3]. Since the subsequence is uniformly bounded in their norms the functions are equicontinuous and therefore converge uniformly on a compact set. The same is true for every bounded sequence of functionals on  $X$  as we are concerned only with their values on  $Y$ . It follows that  $U(x)$  is completely continuous.

Gelfand has proved the following corollary for the case  $\Gamma = \bar{X}$  [12, p. 268]. It should be pointed out that the corollary is not true for non-separable  $X$  as was stated by Gelfand. In his argument he falsely assumed that the functionals of a weakly convergent sequence of functionals of a closed linear subspace of  $X$  could be extended so that the sequence converged weakly on  $X$  (see 7.5).

**3.2. COROLLARY.** *A necessary and sufficient condition for a subset  $S$  of a separable Banach space  $X$  to be conditionally compact is that all weakly convergent sequences of functionals of  $\Gamma$  on  $X$  converge uniformly on  $S$ .*

Every bounded sequence of functionals on  $X$  contains a weakly convergent subsequence [1, p. 123, Theorem 3]. By hypothesis this sequence converges uniformly on  $S$  and hence the transformation  $U(x) = x(x)$  on  $\Gamma$  to  $M_S$  is completely continuous. By Theorem 3.1  $S$  is conditionally compact. The necessity argument is similar to that used in Theorem 3.1.

The following lemma will permit us to prove that the corollary can not be extended to non-separable spaces even if, in its statement,  $\Gamma$  is replaced by  $\bar{X}$ . We now suppose  $T$  to be the class of all positive integers  $t$  and  $\mathcal{T}$  the family of all subsets  $\tau$  of  $T$ .

**3.3. LEMMA.** *If  $\beta^n(\tau)$  are bounded and finitely additive set functions on  $\mathcal{T}$  to real numbers, and if  $\lim_n \beta^n(\tau) = 0$  for all  $\tau \in \mathcal{T}$ , then  $\lim_n \sum_i |\beta^n(t_i)| = 0$ .*

Suppose the lemma to be false. Then there exists an  $\epsilon > 0$  such that  $\limsup_n \sum_i |\beta^n(t_i)| > \epsilon$ . Now  $\beta^n(t) \rightarrow 0$  as  $n \rightarrow \infty$ . Hence we can choose two increasing sequences of integers  $n_i, N_i$  such that  $\sum_{N_i+1}^{N_{i+1}-1} |\beta^{n_i}(t)| \geq \epsilon$  and  $\sum_{N_i+1}^{N_{i+1}-1} |\beta^{n_i}(t)| + \sum_{N_{i+1}}^{\infty} |\beta^{n_i}(t)| < \epsilon/8$ . Let us consider for the moment as a primary block some subset  $\tau_i$  of  $N_i \leq t < N_{i+1}$  for which  $|\beta^{n_i}(\tau_i)| > \epsilon/2$ .  $\sum \tau_i$  is then divided into a denumerable set of disjoint blocks. Since a denumerable set has an infinite number of disjoint denumerable subsets and since  $\beta^n(\tau)$  is bounded, there will exist a denumerable subset of blocks  $\pi_1$  such that, on any of its subsets  $\pi$ ,  $|\beta^{n_1}(\pi)| \leq \epsilon/8$ . The same argument gives a denumerable subset  $\pi_2$  of  $\pi_1$  such that, on any of its subsets  $\pi$ ,  $|\beta^{n_2}(\pi)| \leq \epsilon/8$ . Likewise we can find a denumerable subset  $\pi_p$  of  $\pi_{p-1}$  such that on any of its subsets  $\pi$ ,  $|\beta^{n_p}(\pi)| \leq \epsilon/8$ . Clearly  $\tau_p \notin \pi_p$ . Let  $\pi_0$  consist of the  $n$ th block of  $\pi_n$  for all  $n$ . If  $\pi_0$  contains the block  $\tau_k$ , then there exists  $q_k \geq k$  such that

$$\beta^{n_k}(\pi_0) = \sum_{i=1}^{q_k} \beta^{n_k}(\tau_i \cdot \pi_0) + \beta^{n_k}(\pi_k \cdot \pi_0).$$



Therefore

$$|\beta^{n_0}(\pi_0)| \geq |\beta^{n_0}(\tau_k)| - e/8 - e/8 \geq e/4.$$

Since  $\pi_0$  contains a denumerable number of such blocks,  $\beta^{n_0}(\pi_0)$  does not approach 0, which is contrary to our hypothesis.

As Hildebrandt [14] has shown, to every  $\bar{x} \in \bar{m}$  there corresponds a unique additive bounded set function  $\beta(\tau)$  on all sets of integers such that for all  $x \in m$ ,  $\bar{x}(x) = \int \tau x(t) d\beta$ . If  $\bar{x}_n(x) = \int \tau x(t) d\beta^n$  converge weakly to zero on  $m$ , then  $\beta^n(\tau) \rightarrow 0$  for all  $\tau \in \mathcal{C}$ . We therefore have the following

3.4. COROLLARY. If  $\bar{x}_n(x) = \int \tau x(t) d\beta^n$  converge weakly to zero on  $m$ , then  $\sum_i |\beta^n(t_i)| \rightarrow 0$  as  $n \rightarrow \infty$ .

3.5. EXAMPLE. Let  $S$  be the set of unit vectors  $x_p$  in  $m$ . If  $\bar{x}_n$  converges weakly to  $\bar{x}_0$ , then  $\bar{y}_n = \bar{x}_n - \bar{x}_0$  converges weakly to zero. As above  $\bar{y}_n(x) = \int \tau x(t) d\beta^n$  and  $\sum_i |\beta^n(t_i)| \rightarrow 0$  as  $n \rightarrow \infty$ . Therefore  $\bar{y}_n(x_p) = \beta^n(p) \rightarrow 0$  uniformly in  $p$ . In other words,  $\bar{x}_n(x_p) \rightarrow \bar{x}_0(x_p)$  uniformly in  $S$ .

3.6. THEOREM. If  $U$  is additive and homogeneous on  $X$  to  $Y$  and  $\bar{U}$  is completely continuous on  $\Gamma$  to  $\bar{X}$ , then  $U$  is completely continuous on  $X$  to  $Y$ .

As Dunford [7, p. 317, Theorem 18] has shown, this hypothesis is sufficient to make  $U$  a linear transformation on  $X$  to  $Y$ . Let  $S$  be the image under  $U$  of  $X_1$ , the unit sphere of  $X$ .  $\bar{y}(U(x))$  is then a linear transformation on  $\Gamma$  to  $M_{\bar{X}}$ . Given any sequence  $\{y_n\}$  in the unit sphere of  $\Gamma$ , there exists a subsequence  $n'$  such that  $\bar{U}(y_{n'})$  converges in  $\bar{X}$ . Hence  $\bar{y}_{n'}(U(x)) = \bar{U}(y_{n'})(x)$  converges in  $M_{\bar{X}}$ . We can now apply Theorem 3.1 with  $S = U(X_1)$ .  $S$  is conditionally compact and hence  $U$  is completely continuous.

We will now give a second characterization of conditionally compact sets in a Banach space<sup>(6)</sup>.  $\Pi$  will be a general range of elements  $\pi$  transitive and compositive with respect to the relation  $\geq$ .  $U_\pi$  will be a set of completely continuous transformations on  $X$  to  $X$  defined on  $\Pi$  with the properties: (1) For every  $x \in X$ ,  $\lim_\pi U_\pi(x)$  exists and is equal to  $x$ . (2) There exists a positive number  $M$  such that  $\|U_\pi\| \leq M$  for all  $\pi \in \Pi$ . When the  $U_\pi$  are in addition degenerate<sup>(7)</sup>, such a class of transformations is called a generalized base of  $X$ .

3.7. THEOREM. Necessary and sufficient conditions that a set  $S \subset X$  be conditionally compact are

- (1) L.U.B.  $\{\|x\| \mid x \in S\} < \infty$ ,
- (2)  $\lim_\pi \|U_\pi(x) - x\| = 0$  uniformly in  $S$ .

If we suppose  $S$  to be conditionally compact, then given  $\epsilon > 0$ , there exists

<sup>(6)</sup> Dr. T. H. Hildebrandt suggested Theorem 3.7 as a generalization of the author's application of it to  $L^p$ .

<sup>(7)</sup> A degenerate transformation on  $X$  to  $Y$  is a linear transformation on  $X$  to a finite dimensional subspace of  $Y$ .



$x_1, x_2, \dots, x_n \in X$  such that for any  $x \in S$  there is a  $k$  for which  $\|x - x_k\| < \epsilon$ . For the set  $x_1, x_2, \dots, x_n$  there exists a  $\pi_\epsilon$  such that if  $\pi \geq \pi_\epsilon$ , then  $\|U_\pi(x_k) - x_k\| \leq \epsilon$ . Therefore if  $x \in S$  and  $\pi \geq \pi_\epsilon$ ,

$$\|U_\pi(x) - x\| \leq \|U_\pi(x - x_k)\| + \|U_\pi(x_k) - x_k\| + \|x_k - x\| \leq \epsilon(2 + M),$$

which proves the necessity.

The sufficiency argument is as follows: Given  $\epsilon > 0$ , there exists  $\pi_\epsilon$  such that  $\|U_{\pi_\epsilon}(x) - x\| < \epsilon/3$  for all  $x \in S$ . As  $U_{\pi_\epsilon}$  is completely continuous and as L.U.B.  $\{\|x\| \mid x \in S\} < \infty$ , it follows that there exist  $x_1, x_2, \dots, x_n \in S$  such that for any  $x \in S$  there is a  $k$  for which  $\|U_{\pi_\epsilon}(x) - U_{\pi_\epsilon}(x_k)\| \leq \epsilon/3$ . Therefore

$$\|x - x_k\| \leq \|x - U_{\pi_\epsilon}(x)\| + \|U_{\pi_\epsilon}(x - x_k)\| + \|U_{\pi_\epsilon}(x_k) - x_k\| \leq \epsilon.$$

$S$  is therefore totally bounded or, its equivalent, conditionally compact.

Theorem 3.7 gives a characterization of conditionally compact sets in  $X$  which contains as a special case that given by Kolmogoroff [18], Tamarkin [26], and Tulajkov [27] for  $L^p(\alpha)$  where  $T = (0, 1)$ ,  $\alpha$  is the Lebesgue measure, and  $1 \leq p < \infty$ . In this case  $\Pi$  is the set of integers and

$$U_n(\phi) = \frac{n}{2} \int_{t-1/n}^{t+1/n} \phi(s) ds \quad (\phi \in L^p).$$

For  $1 < p < \infty$ ,  $[U_n(\phi) \mid \|\phi\| \leq 1]$  is uniformly bounded and equi-absolutely continuous, and therefore is conditionally compact in  $L^p$ . For  $p=1$ ,  $[U_n(\phi) \mid \|\phi\| \leq 1]$  is of uniform bounded variation, and therefore is conditionally compact in  $L$ . Finally  $\|U_n\| \leq 1$  for  $1 \leq p < \infty$ . The conclusions of the theorem are consequently valid.

Theorem 3.7 can also be applied to the spaces  $L^p(\alpha)$  ( $1 \leq p \leq \infty$ ) where  $T$  is an abstract class of elements. For  $1 \leq p < \infty$ , let  $\pi$  be of type 1 and contain only a finite number of disjoint measurable sets  $(\tau_1, \tau_2, \dots, \tau_n)$ .  $\chi_\tau$  will denote the characteristic function of the set  $\tau$ . Finally we define  $U_\pi$  on  $L^p(\alpha)$  to  $L^p(\alpha)$  to be

$$U_\pi(\phi) = \sum_{\tau} \frac{\int_{\tau} \phi}{|\tau_i|} \chi_{\tau_i}.$$

For  $p = \infty$ , let  $\pi$  be of type 2 and contain the disjoint measurable sets  $(\tau_1, \tau_2, \dots, \tau_n)$ . Then  $U_\pi$  on  $L^\infty(\alpha)$  to  $L^\infty(\alpha)$  will be

$$U_\pi(\phi) = \sum_{\tau} \frac{\int_{\sigma_i} \phi}{|\sigma_i|} \chi_{\sigma_i}$$

where  $\sigma_i$  is some set of finite measure contained in  $\tau_i$ . The  $U_\pi$  clearly define a generalized base for  $L^p(\alpha)$ .

4. **Linear transformations.** In the following discussion for  $1 \leq p < \infty$ ,  $\pi$  will be of type 1; while for  $p = \infty$ ,  $\pi$  will be of type 2.  $L^p$  ( $1 \leq p < \infty$ ) will be of the

space  $L^p(\alpha)$ .  $L^\infty$  will be either (a) the space of bounded  $\mathfrak{G}$ -measurable functions, or (b) the space of  $\mathfrak{G}$ -measurable functions essentially bounded relative to a measure  $\alpha$  with  $x(\tau) \in V^1(X, \Gamma)$  vanishing on null sets in the latter case.  $1/p + 1/q = 1$ .

4.1. THEOREM. The general form of the linear transformation  $U(\phi)$  on  $L^p$  ( $1 \leq p \leq \infty$ ) to  $X$  is

$$U(\phi) = \int \phi dx$$

where  $x \in V^q(X, \Gamma)$  and  $\|U\| = \|x\|$ .

If  $x \in V^q(X, \Gamma)$  then it follows from Theorems 2.3, 2.4, 2.7, and 2.8 that  $\int \phi dx$  is a linear transformation on  $L^p$  to  $X$  with  $\|U\| = \|x\|$ .

To demonstrate the converse, let  $\chi_r(t)$  be the characteristic function of  $\tau \in \mathfrak{G}$  for  $|\tau| < \infty$ . We define

$$x(\tau) = U[\chi_r(t)].$$

$x(\tau)$  is obviously additive on sets  $\tau$  of finite measure. For  $p=1$ ,

$$\begin{aligned} \|U\| &= \text{L.U.B. } [|\int x[U(\phi)]| = |\bar{U}(x)[\phi]| \mid \|x\| = 1, \|\phi\| = 1] \\ &= \text{L.U.B. } \left[ \left| \frac{\bar{U}(x)[\chi_r]}{|\tau|} \right| \mid \|x\| = 1, \tau \in \mathfrak{G}, |\tau| < \infty \right] \\ &= \text{L.U.B. } \left[ \frac{\|x(\tau)\|}{|\tau|} \mid \tau \in \mathfrak{G}, |\tau| < \infty \right]. \end{aligned}$$

Therefore  $x \in V^\infty(X)$  and  $\|U\| = \|x\|$ . For  $1 < p < \infty$ , we define, for a given  $\pi$  and  $\{a_i\} \in l^p$ ,  $\phi(t) = a_i |\tau_i|^{-1/p}$  when  $t \in \tau_i$ . Then

$$\|\phi\| = \left\{ \int |\phi|^p d\alpha \right\}^{1/p} = \left\{ \sum |a_i|^p \right\}^{1/p} = \|a\|,$$

$$U(\phi) = \sum_i \phi(t_i) x(\tau_i) = \sum_i \frac{a_i x(\tau_i)}{|\tau_i|^{1/p}}, \quad (t \in \tau_i).$$

Finally

$$\begin{aligned} \|U\| &= \text{L.U.B. } \left[ \left| \sum_i \frac{a_i x(\tau_i)}{|\tau_i|^{1/p}} \right| \mid \|x\| = 1, x \in \Gamma, \pi, a \in l^p, \|a\| = 1 \right] \\ &= \text{L.U.B. } \left[ \left\{ \sum_i \frac{|x(\tau_i)|^q}{|\tau_i|^{q-1}} \right\}^{1/q} \mid \pi, x \in \Gamma, \|x\| = 1 \right] = \|x\|. \end{aligned}$$

Again  $x \in V^q(X, \Gamma)$  and  $\|U\| = \|x\|$ . For  $p = \infty$ ,

$$\begin{aligned}\|U\| &= \text{L.U.B.} \left[ \left| \sum_{\tau} \phi(t_i) \hat{x}[x(\tau_i)] \right| \mid \hat{x} \in \Gamma, \|\hat{x}\| = 1, \pi, |\phi(t_i)| \leq 1 \right] \\ &= \text{L.U.B.} \left[ \sum_{\tau} |\hat{x}[x(\tau_i)]| \mid \hat{x} \in \Gamma, \|\hat{x}\| = 1, \pi \right] = \|\hat{x}\|,\end{aligned}$$

$\hat{x} \in V^1(X, \Gamma)$ . To each  $\phi(t) \in L^p$  and  $\pi$  we associate the multiple valued function  $\phi_*(t) = \phi(\tau_i)$  for  $t \in \tau_i$  where  $t_i \in \tau_i$ . Then  $\phi = \lim_{\pi} \phi_*$  in  $L^p$  and by Lemma 2.2 and Theorems 2.3 and 2.7

$$U(\phi) = \lim_{\pi} U(\phi_*) = \int \phi d\hat{x}.$$

For  $p = \infty$  Theorem 4.1 has been demonstrated by Gowurin [13, pp. 265-266].

4.2. COROLLARY.  $V^q(X, \Gamma)$  is equivalent to  $V^q(X, \bar{X})$ ,  $1 \leq p \leq \infty$ .

By Theorem 4.1,  $V^q(X, \Gamma)$  is equivalent to the space of all linear transformations on  $L^p$  to  $X$  for all  $\Gamma$ .

4.3. THEOREM. The general form of the completely continuous transformation  $U(\phi)$  on  $L^p$  ( $1 \leq p \leq \infty$ ) to  $X$  is

$$U(\phi) = \int \phi d\hat{x}$$

where  $\hat{x} \in V^q(X, \Gamma)$  and  $\|U\| = \|\hat{x}\|$ .

This is an immediate consequence of Theorems 3.6 and 4.1.

4.4. COROLLARY.  $V^q_c(X, \Gamma)$  is equivalent to  $V^q_c(X, \bar{X})$ ,  $1 \leq q \leq \infty$ .

4.5. THEOREM. The general form of the linear transformation  $U(x)$  on  $X$  to  $V^q(R)$  ( $1 < q < \infty$ ) where  $T = \sum_{i=1}^{\infty} \tau_i$  ( $|\tau_i| < \infty$ ) is

$$U(x) = \hat{x}(\tau)[x]$$

where  $\hat{x} \in V^q(\bar{X}, X)$  and  $\|U\| = \|\hat{x}\|^{(*)}$ .

It is clear that  $\hat{x} \in V^q(\bar{X}, X)$  defines such a transformation and that  $\|U\| = \|\hat{x}\|$ . Conversely, if  $U$  is linear on  $X$  to  $V^q(R)$ , then its adjoint  $\bar{U}$  defines a transformation on  $V^p(R)$  or its equivalent  $L^p(\alpha)$  to  $\bar{X}$ . By Theorem 4.1

$$\bar{U}(\phi) = \int \phi d\hat{x}$$

where  $\phi \in L^p(\alpha)$ ,  $\hat{x} \in V^q(\bar{X}, X)$ , and  $\|\bar{U}\| = \|\hat{x}\|$ . Since  $\phi[U(x)] = \int \phi d\hat{x}(\hat{x})$  for every  $\phi \in L^p(\alpha)$ , it follows that  $U(x) = \hat{x}(\tau)[x]$  and  $\|U\| = \|\bar{U}\| = \|\hat{x}\|$ .

(\*) Compare with Kantorovitch and Vulich [17, pp. 133-135].

4.6. THEOREM. *The general form of the completely continuous transformation  $U(x)$  on  $X$  to  $V^q(R)$  ( $1 < q < \infty$ ) where  $T = \sum_{i=1}^{\infty} \tau_i$  ( $|\tau_i| < \infty$ ) is*

$$U(x) = \hat{x}(\tau)[x]$$

where  $\hat{x} \in V^q(\overline{X}, X)$  and  $\|U\| = \|\hat{x}\|$ .

The argument used in Theorem 4.5 applies here if we note that  $U$  is necessarily completely continuous [1, p. 101, Theorem 4]. The reference is now made to Theorem 4.3 instead of Theorem 4.1.

If  $T = (0, 1)$  and  $\alpha$  is the Lebesgue measure, then the general form of the linear (or completely continuous) transformation on  $X$  to  $L^q$  ( $1 < q < \infty$ ) is

$$U(x) = \frac{d}{dt} \hat{x}(I_0^t) \cdot [x]$$

where  $\hat{x} \in V^q(\overline{X}, X)$  (or  $V^q_c(\overline{X}, X)$ ),  $\|U\| = \|\hat{x}\|$  and  $I_0^t = (0, t)$ . This is a slightly stronger result than that found by Bochner and Taylor [5, pp. 941-944, Theorems 8.1 and 8.4].

A linear transformation on  $L^p$  to  $L^{p'}$ , where  $\alpha$  is the Lebesgue measure on  $(0, 1)$ , is characterized by a function  $K(s, \tau)$  for which

$$\frac{d}{dt} \int_0^1 \psi(s) K(s, I_0^t) ds \in L^q$$

for every  $\psi \in L^{q'}$  ( $I_0^t = (0, t)$ ). If in addition the transformation is completely continuous, then  $K(s, \tau)$  also satisfies the condition

$$\lim_{h \rightarrow 0} \int_0^1 \left| \frac{d}{dt} \int_0^1 \psi(s) K(s, I_0^t) ds \right|_{t+h} - \frac{d}{dt} \int_0^1 \psi(s) K(s, I_0^t) ds \Big|_t^q dt = 0$$

uniformly for all  $\psi \in L^{q'}$ .

We leave the proof of the following theorems to the reader. Except for the space  $c_0$ , the argument is a special case of the above. Gelfand has discussed the space  $c$  [12, pp. 272-275]. It is convenient to denote the space  $c_0$  by the symbol  $l^\infty$ .

4.7. THEOREM. *The general form of the linear [or completely continuous] transformation  $U(a)$  on  $l^p$  ( $1 \leq p \leq \infty$ ) to  $X$  is*

$$U(a) = \sum a_i x_i$$

where  $\hat{x} \in v^q(X, \Gamma)$  [or  $\hat{x} \in v^q_c(X, \Gamma)$ ] and  $\|U\| = \|\hat{x}\|$ .

4.8. COROLLARY.  $v^q(X, \Gamma)$  [or  $v^q_c(X, \Gamma)$ ] is equivalent to  $v^q(X, \overline{X})$  [or  $v^q_c(X, \overline{X})$ ] ( $1 \leq q \leq \infty$ ).

4.9. COROLLARY. *If  $X$  is either weakly complete or a separable conjugate space, then any linear transformation on  $c_0$  to  $X$  is completely continuous.*

$\hat{x} \in v^1(X, \bar{X})$  implies that  $\hat{x} \in v^1_c(X, \bar{X})$  according to a theorem of Orlicz [21, pp. 244-247] and Theorem 1.8. The conclusion follows from Theorem 4.7.

4.10. THEOREM. *The general form of the linear [or completely continuous] transformation  $U(x)$  on  $X$  to  $l^q$  ( $1 \leq q < \infty$ ) is*

$$U(x) = \{\hat{x}_i(x)\}$$

where  $\hat{x} \in v^q(\bar{X}, X)$  [or  $\hat{x} \in v^q_c(\bar{X}, X)$ ] and  $\|U\| = \|\hat{x}\|$ .

4.11. COROLLARY. *If  $X$  is either weakly complete or a separable conjugate space, then any linear transformation on  $X$  to  $l^1$  is completely continuous.*

We conclude this section with some considerations about linear transformations on  $C$  to  $X$ . It follows from Theorem 2.10 that  $U(\phi) = \int \phi d\hat{x}$  where  $\phi \in C$  and  $\hat{x} \in V(X)$  is a linear transformation on  $C$  to  $X$  with  $\|U\| = \|\hat{x}\|$ . Gelfand [12, p. 283] has shown that the general form of a completely continuous transformation on  $C$  to  $X$  is

$$U(\phi) = \int \phi d\hat{x}$$

where  $\hat{x} \in V_c(X)$  and  $\|\hat{x}\| = \|U\|$ . When  $X$  is weakly complete, Gelfand has shown this to be the general form of the linear transformation on  $C$  to  $X$  where now  $\hat{x} \in V(X)$ . It might be added that Gelfand's method will show this to be true for all conjugate Banach spaces  $X$ .

It is easy to give an example of a linear transformation on  $C$  to  $X$  which does not have this general form. Let  $U$  be the identity transformation on  $C$  to  $C$  and suppose that it does have this form. Then  $\phi(s) = U(\phi) = \int \phi(t) d\psi_t(s)$ . As this holds for all  $\phi \in C$ ,  $\psi_t(s) = c$  ( $s > t$ ) and  $\psi_t(s) = 1 + c$  ( $s < t$ ), which is contrary to  $\psi_t(s) \in C$  for fixed  $t$ . Because of the above remark, this again shows that  $C$  is not a conjugate space.

The following theorem gives a characterization for linear transformations on  $C$  to  $X$ :

4.12. THEOREM. *A necessary and sufficient condition that  $U$  be a linear transformation on  $C$  to  $X$  is that there exist a sequence of step functions  $\hat{x}_n \in V(X)$  such that  $\lim_{n \rightarrow \infty} \|\hat{x}_n\| = \|U\|$  and*

$$U(\phi) = \lim_n \int \phi d\hat{x}_n.$$

Making use of the Bernstein polynomials

$$\phi(t) = \lim_{n \rightarrow \infty} \sum_{r=0}^n C_r^n \phi(r/n) t^r (1-t)^{n-r}$$

in  $C$ . If we apply a device due to Hildebrandt and Schoenberg [15, p. 318], then

$$U(\phi) = \lim_{n \rightarrow \infty} \sum_{r=0}^n \phi(r/n) U[C_r^n(1-t)^{n-r}] = \lim_{n \rightarrow \infty} U_n(\phi)$$

where  $U_n(\phi) = \int \phi dx_n$ ,  $x_n(t)$  being defined to have the jump  $U[C_r^n(1-t)^{n-r}]$  at  $r/n$  and to be constant elsewhere. Now

$$\|U_n(\phi)\| = \|U[\sum C_r^n \phi(r/n) t^r (1-t)^{n-r}]\| \leq \|U\| \cdot \|\phi\|.$$

Therefore  $\|\dot{x}_n\| = \|U_n\| \leq \|U\|$ . In general, however,  $\liminf_{n \rightarrow \infty} \|U_n\| \geq \|U\|$  so that  $\lim_{n \rightarrow \infty} \|\dot{x}_n\| = \|U\|$ . Since  $\dot{x}_n$  does define a linear transformation, the sufficiency argument is obvious [1, p. 80, Theorem 5].

**5. Linear transformations on  $L$ .** In this section we obtain representations by means of a kernel of the general completely continuous transformation and weakly completely continuous separable transformation on  $L$  to an arbitrary Banach space  $X$ . By means of this result and a theorem due to Dunford and Pettis [9], we show that  $U^2$  is completely continuous whenever  $U$  on  $L$  to  $L$  is weakly completely continuous. Special cases of Theorems 5.3 and 5.4 have been proved by Gelfand [12]. More recently Dunford and Pettis [9] have obtained special cases of Theorems 5.3, 5.4, and 5.5.

In this and the following section,  $L^p$  will be the space  $L^p(\alpha)$  where  $T$  is the sum of a denumerable number of sets  $\tau \in \mathcal{T}$  of finite measure.  $\pi$  will be defined as at the end of §3.  $x(t)$  on  $T$  to  $X$  will be said to be weakly measurable if  $\dot{x}(x(t))$  is measurable for all  $\dot{x} \in \bar{X}$ . We define  $B^w(X)$  to be the class of weakly measurable point functions  $x(t)$  on  $T$  to  $X$  whose values are essentially contained in a separable conditionally weakly compact subspace of  $X$ . With norm

$$\|\dot{x}\| = \text{ess. L.U.B.}_t [\|x(t)\|],$$

$B^w(X)$  is a Banach space. The set of functions  $x(t) \in B^w(X)$  which take on a.e. a conditionally compact set of values will comprise the subspace  $B^w_c(X)$  of  $B^w(X)$ .

Integration with respect to a real valued measure function  $\alpha$  will be realized by means of the Birkhoff integral [3]. Since  $x(t)$  for  $\dot{x} \in B^w(X)$  is a.e. contained in a separable subspace of  $X$ ,  $x(t)$  is integrable on all sets  $\tau \in \mathcal{T}$  of finite measure [22, Theorems 1.1 and 5.3, Corollary 5.11].

**5.1. LEMMA.** If  $\dot{x} \in B^w(X)$ , then

$$\|\dot{x}\| = \text{L.U.B.} \left[ \frac{\left\| \int_{\tau} x(t) d\alpha \right\|}{|\tau|} \mid 0 < |\tau| < \infty, \tau \in \mathcal{T} \right].$$

Since  $x(t)$  is essentially contained in a separable subspace  $X'$  of  $X$ , it is clear that there exists a denumerable set of linear functionals  $\{\dot{x}_n\} \subset \bar{X}$  each



of norm one such that L.U.B.  $[|\hat{x}_n(x)| = \|x\| \mid n, x \in X']$ . Therefore

$$\begin{aligned} A &= \text{L.U.B.} \left[ \frac{\left| \int_{\tau} x(t) d\alpha \right|}{|\tau|} \mid 0 < |\tau| < \infty, \tau \in \mathfrak{G} \right] \\ &= \text{L.U.B.} \left[ \frac{\left| \int_{\tau} \hat{x}_n(x(t)) d\alpha \right|}{|\tau|} \mid 0 < |\tau| < \infty, \tau \in \mathfrak{G}, n \right]. \end{aligned}$$

Let  $\tau_n = [t \mid |\hat{x}_n(x(t))| > A]$ . Clearly  $|\tau_n| = 0$ .  $\tau_0 = [t \mid \|x(t)\| > A, x(t) \in X']$  is contained in  $\sum \tau_n$  and hence is of measure zero. On the other hand for every  $\epsilon > 0$ , there exists an  $n$  such that  $\text{ess. L.U.B.}_t |\hat{x}_n(x(t))| > A - \epsilon$ . It follows that  $\text{ess. L.U.B.}_t \|x(t)\| = A$ .

**5.2. THEOREM.**  $V^{\infty}_e(X)$  is equivalent to  $B^{\infty}_e(X)$ . The equivalence is defined by  $U(x(t)) = x(\tau) = \int x(t) d\alpha$  on  $B^{\infty}_e(X)$  to  $V^{\infty}_e(X)$ .

It follows from Lemma 5.1 that  $U$  is an isometric transformation. By the definition of the Birkhoff integral, given any null set  $\tau_0$ ,  $S = [x(\tau)/|\tau| \mid 0 < |\tau| < \infty, \tau \in \mathfrak{G}]$  is contained in the convex extension of  $[x(t) \mid t \in T - \tau_0]$ . Therefore  $S$  is conditionally compact. By Theorem 1.4,

$$x(\tau) = U(x(t)) \in V^{\infty}_e(X).$$

We next prove the converse. Let  $x(\tau) \in V^{\infty}_e(X)$ .  $\hat{x}(x(\tau))$  is then a completely additive set function on all measurable subsets of any set of finite measure. By the Radon-Nikodym theorem there exists  $f_x(t) \in B^{\infty}(R)$  such that  $\hat{x}(x(\tau)) = \int_{\tau} f_x(t) d\alpha$  for all  $\tau \in \mathfrak{G}$  of finite measure. As above this defines an isometric transformation  $V$  on  $B^{\infty}(R)$  to  $V^{\infty}(R)$ . By definition,  $[\hat{x}(x(\tau)) \mid \hat{x} \in \bar{X}, \|\hat{x}\| \leq 1]$  is conditionally compact in  $V^{\infty}(R)$ . Therefore  $P = [f_x(t) \mid \hat{x} \in \bar{X}, \|\hat{x}\| \leq 1]$  is conditionally compact in  $B^{\infty}(R)$ . Defining  $U_x$  as in §3, it follows from Theorem 3.7 that  $\lim_{\tau} U_x(f_x(t)) = f_x(t)$  uniformly in  $P$  in the topology of  $B^{\infty}(R)$ .

$$V(U_x(f_x(t))) = \sum_{\tau} \frac{\hat{x}[x(\sigma_i)]}{|\sigma_i|} \cdot |\tau \cdot \tau_i|.$$

Then

$$\lim_{\tau} V(U_x(f_x(t))) = V(f_x(t))$$

uniformly in  $P$ . This implies

$$(1) \quad \lim_{\tau} \sum_{\tau} \frac{x(\sigma_i)}{|\sigma_i|} \cdot |\tau \cdot \tau_i| = x(\tau)$$

in  $V^{\infty}_e(X)$ . Define  $x_{\tau}(t) = x(\sigma_i)/|\sigma_i|$  for  $t \in \tau_i$ ; and  $x_{\tau}(\tau) = \int x_{\tau}(t) d\alpha$ . Then  $x_{\tau}(t) \in B^{\infty}_e(X)$ ;

$$x_\tau(\tau) = \sum_i \frac{x(\sigma_i)}{|\sigma_i|} |\tau \cdot \tau_i| = U(x_\tau(t)).$$

By (1), given  $\epsilon > 0$ , there exists a  $\pi_\epsilon$  such that for  $\pi_1, \pi_2 \geq \pi_\epsilon$ ,  $\|x_{\pi_1}(\tau) - x_{\pi_2}(\tau)\|_{V^\infty(X)} \leq \epsilon$ . According to Lemma 5.1, ess. L.U.B.,  $\|x_{\pi_1}(t) - x_{\pi_2}(t)\| \leq \epsilon$ . Since  $B^\infty_c(X)$  is a Banach space, it follows from Lemma 2.2 that there exists an  $x(t) \in B^\infty_c(X)$  such that  $\lim_\tau x_\tau(t) = x(t)$  in  $B^\infty_c(X)$ . Then  $x(\tau) = \lim_\tau U(x_\tau(t)) = U(x(t))$ .

**5.3. THEOREM.** *The general form of the completely continuous transformation  $U$  on  $L$  to  $X$  is*

$$U(\phi) = \int \phi(t)x(t)d\alpha$$

where  $x \in B^\infty_c(X)$  and  $\|U\| = \|x\|$ .

According to Theorem 4.3,  $U(\phi) = \int \phi(t)dx$  where  $x(\tau) \in V^\infty_c(X)$  and  $\|U\| = \|x(\tau)\|_{V^\infty_c(X)}$ . By Theorem 5.2 there exists an  $x(t) \in B^\infty_c(X)$  such that  $x(\tau) = \int_\tau x(t)d\alpha$  for all  $\tau \in \mathcal{T}$  of finite measure and  $\|x(t)\|_{B^\infty_c(X)} = \|x(\tau)\|_{V^\infty_c(X)} = \|U\|$ . Further by Theorem 2.5,  $\int_\tau \phi(t)dx = \int_\tau \phi(t)x(t)d\alpha$  for all  $\tau \in \mathcal{T}$  of finite measure. Now  $T = \sum \tau_i$  where  $|\tau_i| < \infty$  and  $\tau_i \cdot \tau_j = 0$  if  $i \neq j$ . Given  $\epsilon > 0$ , one can obtain an unconditionally convergent sum of the type  $\sum_i \phi(t_i)x(t_i)|\tau_j^i|$  ( $(\tau_j^i)$  is a subdivision of  $\tau_i$ ,  $t_i \in \tau_j^i$ ) which approximates  $\int_\tau \phi(t)x(t)d\alpha$  to within  $\epsilon/2^i$  and such that each of its finite partial sums is within  $\epsilon/2^i$  of some  $\int_\tau \phi(t)x(t)d\alpha$  for  $\tau \subset \tau_i$ . Since by Theorem 2.4,  $\|\int_\tau \phi(t)x(t)d\alpha\| \leq \|x\| \int_\tau |\phi(t)|d\alpha$ , it follows that the resulting subdivision of  $T$  furnishes an unconditionally convergent sum which approximates  $\int_T \phi(t)dx$  to within  $\epsilon$ .  $\int_T \phi(t)x(t)d\alpha$  therefore exists and is equal to  $\int_T \phi(t)dx$ .

**5.4. THEOREM.** *The general form of the weakly completely continuous separable transformation on  $L$  to  $X$  is*

$$U(\phi) = \int \phi(t)x(t)d\alpha$$

where  $x \in B^\infty(X)$  and  $\|U\| = \|x\|$ .

According to Theorem 4.1,  $U(\phi) = \int \phi(t)dx$  where  $x(\tau) \in V^\infty(X)$  and  $\|U\| = \|x(\tau)\|_{V^\infty(X)}$ . If  $\chi_\tau(t)$  is the characteristic function of  $\tau \in \mathcal{T}$  for  $|\tau| < \infty$ , then  $x(\tau) = U(\chi_\tau(t))$ . As  $U$  is weakly completely continuous and separable, it is clear that  $S = [x(\tau)/|\tau| \mid 0 < |\tau| < \infty]$  is conditionally weakly compact and is contained in a separable linear closed subspace  $Y$  of  $X$ . Hence there exists a sequence  $\{x_n\} \subset Y$  which when considered as elements of  $\bar{Y}$  are dense in the unit sphere of a determining manifold in  $\bar{Y}$ . By the Radon-Nikodym theorem there exists for each  $x \in \bar{Y}$  an  $f_x(t) \in B^\infty(R)$  such that  $x(x(\tau)) = \int_\tau f_x(t)d\alpha$  for all  $\tau \in \mathcal{T}$  of finite measure. As in Theorem 5.2, let  $\pi$  be a finite subdivision

of  $T$  into disjoint measurable sets  $(\tau_1, \tau_2, \dots, \tau_n)$ , let  $\sigma_i \subset \tau_i$  such that  $0 < |\sigma_i| < \infty$ , and let  $x_{\sigma_i}(t) = x(\sigma_i)/|\sigma_i|$  for  $t \in \tau_i$ . Then for every  $x \in X$ ,  $\lim_{\sigma} x(x_{\sigma}(t)) = f_{\sigma}(t)$  in  $B^w(R)$ . There will therefore exist a set  $\{\pi_n\}$  such that  $|x_i(x_{\pi_n}(t)) - f_{\pi_n}(t)| < 1/n$  on  $T - \sigma_n$  ( $|\sigma_n| = 0$ ) for all  $i \leq n$ . Hence for each  $i$ ,  $x_i(x_{\pi_n}(t)) \rightarrow f_{\pi_n}(t)$  uniformly on  $T - \sigma_0$  where  $\sigma_0 = \sum \sigma_n$  ( $|\sigma_0| = 0$ ). For a given  $t$ ,  $x_{\pi_n}(t)$  is contained in  $S$ . A subsequence will therefore converge weakly to an element of  $Y$  [1, p. 134, Theorem 2]. We arbitrarily define  $x(t)$  to be the weak limit of one such subsequence. Clearly  $x_i(x(t)) = \lim_n x_i(x_{\pi_n}(t)) = f_{\pi_n}(t)$  on  $T - \sigma_0$ . As  $|x_i(x(t))| \leq U$  on  $T - \sigma_0$ , and  $x(t) \in Y$ , it follows that  $\|x(t)\| \leq U$  on  $T - \sigma_0$ . Further  $x_i(x(t))$  is measurable,  $Y$  is separable, and the sequence  $x_i$  is dense in the unit sphere of a determining manifold in  $Y$  if the  $x_i$  are considered as elements of  $Y$ . From this one can easily show that  $x(t)$  is weakly measurable. As  $x(t)$  is contained in the sequential weak closure of  $S$ ,  $\{x(t) | t \in T\}$  is conditionally weakly compact<sup>(9)</sup>. Therefore  $x(t) \in B^w(X)$ . Now  $x_i(x(\tau)) = \int_{\tau} f_{\pi_n}(t) d\alpha = \int_{\tau} x_i(x(t)) d\alpha$  for all  $\tau \in \mathcal{G}$  of finite measure. As  $\{x_i\}$  is total in  $Y$ , it follows that  $x(\tau) = \int_{\tau} x(t) d\alpha$  for all  $\tau$  of finite measure [23, Theorem 5.3]. By Lemma 5.1  $\|x(t)\|_{B^w(X)} = \|x(\tau)\|_{V^w(X)} = \|U\|$ . The remainder of the argument is identical to that used in Theorem 5.3.

We remark that Theorem 5.4 is applicable to any separable linear transformation on  $L$  to a regular Banach space since the unit sphere of a regular space is weakly compact.

In the following theorem and corollary,  $T$  need not be the sum of a denumerable number of sets  $\tau \in \mathcal{G}$  of finite measure.

**5.5 THEOREM.** *If  $U$  is a weakly completely continuous transformation on  $L$  to an arbitrary Banach space  $X$ , then  $U$  takes conditionally weakly compact sets into conditionally compact sets.*

It is sufficient to show that for any conditionally weakly compact sequence  $\{\phi_n\}$ ,  $\{U(\phi_n)\}$  is conditionally compact. The sequence  $\{\phi_n\}$  is contained in a separable subspace  $L'$  of  $L$  essentially defined on a class  $T' \subset T$  which is the sum of a denumerable number of sets of finite measure<sup>(10)</sup>. Let  $U'$  on  $L'$  to  $X$  be identical with  $U$  on  $L'$ . As  $L'$  is separable,  $U'$  is a separable weakly completely continuous transformation on  $L'$  to  $X$ . Theorem 5.4 is applicable, and hence by a theorem due to Dunford and Pettis [9, p. 547, Theorem 4]  $U'$  takes conditionally weakly compact sets into conditionally compact sets. Since  $\{\phi_n\}$  is conditionally weakly compact in  $L'$ , this concludes the proof.

<sup>(9)</sup> W. L. Chmoulyan has shown that the weak sequential closure of a weakly compact subset of a Banach space is itself weakly compact. See Communications de l'Institut des Sciences Mathématiques et Mécaniques de l'Université de Kharkoff et la Société Mathématique de Kharkoff, (4), vol. 14 (1937), pp. 239-242.

<sup>(10)</sup> One can readily obtain this result by employing an argument similar to that used by Dunford [8, p. 644].

5.6. COROLLARY. If  $U$  is weakly completely continuous on  $L$  to  $L$ , then  $U^2$  is completely continuous.

$U$  takes the unit sphere in  $L$  into a conditionally weakly compact subset of  $L$ , and by Theorem 5.5 its iterate takes this subset into a conditionally compact subset of  $L$ . In other words  $U^2$  is completely continuous.

A uniform mean ergodic theorem for weakly completely continuous transformations on  $L$  to  $L$  is easily obtainable by means of Corollary 5.6 and a mean ergodic theorem due to Kakutani [16] and Yosida [28].

6. On completely continuous transformations. In this section we make further application of our study, demonstrating that each completely continuous transformation on any of the spaces  $L^p$ ,  $l^p$  ( $1 \leq p \leq \infty$ ),  $C$ ,  $c_0$ ,  $M_T$  to an arbitrary Banach space  $X$  can be approximated in the norm by degenerate transformations (see footnote on p. 526). The notation is that of §5.

For  $\hat{x} \in V^q(X)$  ( $1 < q < \infty$ ) we define

$$x_\tau(\tau) = \sum_i \frac{x(\tau_i)}{|\tau_i|} \cdot |\tau \cdot \tau_i|;$$

and for  $\hat{x} \in V^\infty(X)$  we define

$$x_\tau(\tau) = \sum_i \frac{x(\sigma_i)}{|\sigma_i|} \cdot |\tau \cdot \tau_i| \quad (\sigma_i < \tau_i, |\sigma_i| < \infty).$$

Clearly  $\hat{x}_\tau \in V^q(X)$ .

6.1. THEOREM. If  $\hat{x} \in V^q(X)$  ( $1 < q \leq \infty$ ), then  $\lim_\tau \|\hat{x}_\tau - \hat{x}\| = 0$ .

By definition, the set  $[\hat{x}(x(\tau)) \mid \|\hat{x}\| \leq 1]$  where  $\hat{x} \in V^q(X)$  is a conditionally compact subset of  $V^q(R)$ . If we use the usual isometric correspondence between  $V^q(R)$  and  $L^q$ , it follows immediately from Theorem 3.7 that

$$\lim_\tau \|\hat{x}_\tau - \hat{x}\| = \lim_\tau \text{L.U.B.} [\|\hat{x}(\hat{x}_\tau) - \hat{x}(\hat{x})\| \mid \|\hat{x}\| \leq 1] = 0.$$

6.2. COROLLARY. If  $U$  is a completely continuous transformation on  $L^p$  to  $X$  ( $1 \leq p < \infty$ ), then

(1)  $U(\phi) = \int \phi d\alpha$  where  $\hat{x} \in V^q(X)$ ,

(2) If  $U_\tau(\phi) = \sum_i f_i \phi d\alpha \cdot x(\tau_i)/|\tau_i|$  ( $1 < p < \infty$ ) or if  $U_\tau(\phi) = \sum_i f_i \phi d\alpha \cdot x(\sigma_i)/|\sigma_i|$  ( $p = 1$ ), then  $\lim_\tau \|U_\tau - U\| = 0$ .

This is a consequence of Theorems 4.3 and 6.1.

For notational convenience, we write  $c_0 = l^\infty$ .

6.3. THEOREM. If  $U$  is a completely continuous transformation on  $l^p$  ( $1 \leq p \leq \infty$ ) to  $X$ , then

(1)  $U(a) = \sum_{i=1}^\infty a_i x_i$  where  $\hat{x} \in V^q(X)$ ,

(2) if  $U_n(a) = \sum_{i=1}^n a_i x_i$ , then  $\lim_n \|U_n - U\| = 0$ .

For  $p = \infty$ , this is a consequence of Theorems 1.7 and 4.7. For  $1 \leq p < \infty$ , the theorem is a special case of Corollary 6.2.

6.4. THEOREM. If  $\hat{x} \in V^1_c(X)$ , then there exists a non-negative  $\beta(\tau) \in V^1(R)$  such that

$$\lim_{\pi} \sum_{\tau} \frac{x(\tau_i)}{\beta(\tau_i)} \beta(\tau \cdot \tau_i) = \hat{x}$$

where  $\pi$  is of type 2.

By the definition of  $V^1_c(X)$ ,  $S = [\hat{x}(x(\tau)) \mid \|x\| \leq 1]$  is a conditionally compact subset of  $V^1(R)$ . Hence there exists the sequence  $\{\hat{x}_n \mid \|\hat{x}_n\| \leq 1, n = 1, 2, \dots\}$  such that  $\hat{x}_n(x(\tau))$  is dense in  $S$ . Let  $\beta_n(\tau)$  be the absolute variation of  $\hat{x}_n(x(\tau))$ . Define  $\beta(\tau) = \sum_{i=1}^{\infty} (1/2^i) \beta_n(\tau)$ . Clearly  $\beta(\tau) \in V^1(R)$  and  $|\beta_n(\tau)| \leq 2^n |\beta(\tau)|$ .  $\hat{x}_n(x(\tau))$  is therefore absolutely continuous with respect to  $\beta(\tau)$ . The class of all elements of  $V^1(R)$  absolutely continuous with respect to  $\beta(\tau)$  form a closed linear space  $AC(\beta)$ . Let  $\pi = (\tau_1, \tau_2, \dots, \tau_n)$  be of type 2 and such that  $\beta(\tau_i) \neq 0$  ( $i = 1, 2, \dots, n$ ). Define

$$U_{\pi}(\gamma(\tau)) = \sum_{\tau} \frac{\gamma(\tau_i)}{\beta(\tau_i)} \beta(\tau \cdot \tau_i)$$

on  $AC(\beta)$  to  $AC(\beta)$ . Then  $\|U_{\pi}\| \leq 1$ . By a theorem due to Bochner [4, pp. 780-783]  $\lim_{\pi} U_{\pi}(\gamma(\tau)) = \gamma(\tau)$  for all  $\gamma(\tau) \in AC(\beta)$ . By Theorem 3.7,  $\lim_{\pi} U_{\pi}(\hat{x}_n(x(\tau))) = \hat{x}_n(x(\tau))$  uniformly in  $n$  and hence

$$\lim_{\pi} \left\| \sum_{\tau} \frac{x(\tau_i)}{\beta(\tau_i)} \beta(\tau \cdot \tau_i) - \hat{x} \right\| = \lim_{\pi} \text{L.U.B.}_n [\|U_{\pi}[\hat{x}_n(x(\tau))] - \hat{x}_n(x(\tau))\|] = 0.$$

6.5. COROLLARY. If  $U$  is a completely continuous transformation on  $M_T$  [or  $L^{\infty}$ ] to  $X$ , then

- (1)  $U(\phi) = \int \phi dx$  where  $\hat{x} \in V^1_c(X)$ ,
- (2) there exists a  $\beta(\tau) \in V^1(R)$  such that if

$$U_{\pi}(\phi) = \sum_{\tau} \frac{\int_{\tau} \phi d\beta}{\beta(\tau_i)} x(\tau_i),$$

then  $\lim_{\pi} \|U_{\pi} - U\| = 0$ .

This is a consequence of Theorems 4.3 and 6.4.

6.6. COROLLARY. If  $U$  is a completely continuous transformation on  $C$  to  $X$ , then  $U$  is approximable in the norm by degenerate transformations.

According to a result of Gelfand's [12, p. 283],  $U(\phi) = \int \phi dx$  where  $\hat{x} \in V_c(X)$  and  $\|\hat{x}\| = \|U\|$  (see end of §2). Now  $\lim_{t \rightarrow t^+} \hat{x}(x(t))$  and  $\lim_{t \rightarrow t^-} \hat{x}(x(t))$  exist by virtue of  $\hat{x}(x(t))$ 's being of bounded variation. Since the values as-

sumed by  $x(t)$  form a conditionally compact set, it follows that  $x(t^+)$  and  $x(t^-)$  are defined for all  $t$ . Let  $\mathfrak{G}$  be the Jordan field of sets  $\tau$  generated by all open intervals and points of  $(0, 1)$ . If  $\tau$  consists of the disjoint sets  $[(a_i^+, b_i^-), \dots, (a_n^+, b_n^-); c_1, \dots, c_m]$ , define

$$x(\tau) = \sum_{i=1}^n [x(b_i^-) - x(a_i^+)] + \sum_{i=1}^m [x(c_i^+) - x(c_i^-)].$$

Clearly  $x(\tau) \in V_c^1(X)$ ,  $U(\phi) = \lim_{\tau} \sum \phi(t_i) x(\tau_i)$  ( $\pi$  of type 2), and  $\|U\| = \|\dot{x}\|$ . The remainder of the argument follows from Theorem 6.4.

We remark that in Theorems 6.1, 6.4 and Corollaries 6.2, 6.5, 6.6 the  $\pi$ -limit can be replaced by a sequential limit.

The problem of approximating completely continuous transformations on  $X$  to certain spaces  $Y$  by degenerate transformations has been investigated by Maddaus [19]. He shows that this is possible whenever there exist degenerate transformations  $V_n$  such that  $\lim_{n \rightarrow \infty} V_n(y) = y$  for all  $y \in Y$ .

Let  $Y$  be a Banach space possessing a generalized base,  $U_{\pi}$  (see §3). Suppose  $U$  is a completely continuous transformation on  $X$  to  $Y$ . Then the set  $S = \{U(x) | x \in X, \|x\| \leq 1\}$  is conditionally compact. By Theorem 3.7  $\lim_{\pi} U_{\pi}(U(x)) = U(x)$  uniformly for all  $x \in X$  for which  $\|x\| \leq 1$ . It follows that  $\lim_{\pi} \|U_{\pi}(U) - U\| = 0$ . As  $U_{\pi}(U)$  is a degenerate transformation, this gives Maddaus's result in a slightly more general form.

**7. On the extension of linear transformations.** If  $U$  is a linear transformation on  $X$  to  $Y$  and  $Z$  contains  $X$  as a proper subspace, then a linear transformation  $U_1$  on  $Z$  to  $Y$  such that  $U(x) = U_1(x)$  for all  $x \in X$  is called an extension of  $U$ . Any Banach space  $Y$  can be imbedded<sup>(11)</sup> in a space of type  $M_T$ <sup>(12)</sup>. We will designate such a space which contains  $Y$  or its image under an equivalence by  $M_T \supset Y$ .

**7.1. THEOREM.** *The general form of the linear transformation  $U$  on  $X$  to  $M_T$  is*

$$U(x) = [\bar{x}_i(x)]$$

where  $\|U\| = \text{L.U.B. } [\|\bar{x}_i\| | i \in T]$ .

For every  $i \in T$  there exists a linear functional  $\bar{\alpha}_i$  such that  $\bar{\alpha}_i(a) = a(i)$ . Let  $\bar{x}_i = \bar{U}(\bar{\alpha}_i)$ . Then  $a(i) = \bar{\alpha}_i[U(x)] = \bar{x}_i(x)$  and

$$\|U\| = \text{L.U.B. } [\|\bar{\alpha}_i[U(x)]\| = |\bar{x}_i(x)| | i \in T, \|x\| \leq 1] = \text{L.U.B. } [\|\bar{x}_i\| | i \in T].$$

**7.2. COROLLARY.** *Any linear transformation  $U$  on  $X$  to  $M_T$  has an extension  $U_1$  on  $Z \supset X$  to  $M_T$  such that  $\|U\| = \|U_1\|$ .*

<sup>(11)</sup> By an imbedding of  $Y$  into a subspace  $Z$  of  $M$  we shall mean that  $Y$  is equivalent [1, p. 180] to  $Z$ .

<sup>(12)</sup> Let  $T = \Gamma_1$ , the unit sphere of some determining manifold in  $\mathfrak{F}$ . Then  $U(y) = \mathfrak{f}(y)$  on  $Y$  to  $M_T$  defines an equivalence between  $Y$  and a subset of  $M_T$ .



By the Hahn-Banach theorem [1, p. 55, Theorem 2],  $\bar{x}_1$  has a norm preserving extension  $\bar{z}_1$  on  $Z$ .  $U_1(z) = [\bar{z}_1(z)]$  is the required extension.

7.3. COROLLARY. *If  $Y$  is isomorphic with  $M_T$ , then any linear transformation  $U$  on  $X$  to  $Y$  has an extension on  $Z$  to  $Y$ .*

As  $Y$  and  $M_T$  are isomorphic [1, p. 180] there exists a biunique and bi-continuous linear transformation  $V$  on  $Y$  into the entire space  $M_T$ .  $VU$  is then a linear transformation on  $X$  to  $M_T$  which by Corollary 7.2 has the extension  $(VU)_1$  on  $Z$  to  $M_T$ . It is clear that  $V^{-1}(VU)_1$  is the required extension on  $Z$  to  $Y$ .

7.4. COROLLARY. *Any linear transformation  $U$  on  $X$  to  $Y$  has an extension on  $Z \supset X$  to  $Y$  if either of the following is true:*

- (1) *There exists a projection transformation<sup>(13)</sup>  $P$  on  $Z$  to  $X$ .*
- (2) *There exists a projection transformation  $P$  on  $M_T \supset Y$  to  $Y$ .*

If (1) holds then  $U_1 = UP$  is the required extension on  $Z$  to  $Y$ . If (2) holds and  $U_1$  is the extension of Corollary 7.2 on  $Z$  to  $M_T \supset Y$ , then  $PU_1$  is the required extension on  $Z$  to  $Y$ .

In view of Corollary 7.4, the existence of projection transformations on spaces  $M_T \supset Y$  to  $Y$  assumes importance in the study of the extension of linear transformations. As yet we can give only negative results in this direction.

Fichtenholtz and Kantorovitch [10, p. 92] have proved that there does not exist a projection transformation on  $M_T$  to  $C$  where  $T = (0, 1)$  and  $C$  is the space of continuous functions on  $(0, 1)$ . Banach and Mazur [2, p. 111] have shown that for a separable space  $Y$  whose conjugate space is not weakly complete there does not exist a projection transformation on  $C$  to any imbedding of  $Y$  in  $C$ . Consequently there exists no projection transformation on  $M_T \supset Y$  to an imbedding of  $Y$  which is contained in an imbedding of  $C$  in  $M_T \supset Y$ .

If there existed a projection transformation on the space  $C_1$  of functions on  $(0, 1)$  having only discontinuities of the first kind to  $C$  then the methods of Gelfand [12, p. 281] would show that the identity transformation on  $C$  to  $C$  could be expressed in the form  $U(\phi) = \int \phi dx$  where  $x \in V(X)$ . The example at the end of §4 shows that this is not the case. Therefore there exists no projection transformation on  $M_T \supset C$  to an imbedding of  $C$  which is contained in an imbedding of  $C_1$  in  $M_T \supset C$ .

7.5. *There exists no projection transformation on  $m$  to  $c$ .*

If there existed a projection transformation  $P$  on  $m$  to  $c$ , then any weakly convergent sequence of linear functionals  $\{a_p\}$  on  $c$  corresponds to a se-

<sup>(13)</sup> A projection transformation  $P$  is a linear transformation with the property that  $P^2 = P$ .

quence of extensions  $\{x_p = P(a_p)\}$  which is weakly convergent on  $m$ . Now  $a_p(a) = a(p+1) - a(p)$  converges weakly to zero on  $c$ . Using the notation of Corollary 3.4, we have  $x_p(x) = \int_T x(t) d\beta^p$  and  $\sum_i |\beta^p(t_i)| \rightarrow 0$ . Since  $x_p(a) = a_p(a) = a(p+1) - a(p)$ , it follows that  $\beta^p(p+1) = 1 - \beta^p(p)$  which is contrary to the above. There can therefore exist no projection transformation on  $m$  to  $c$ .

## REFERENCES

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.
2. S. Banach and S. Mazur, *Zur Theorie der linearen Dimension*, *Studia Mathematica*, vol. 4 (1933), pp. 100-112.
3. Garrett Birkhoff, *Integration in a Banach space*, these Transactions, vol. 38 (1935), pp. 357-378.
4. S. Bochner, *Additive set functions on groups*, *Annals of Mathematics*, (2), vol. 40 (1939), pp. 769-799.
5. S. Bochner and A. E. Taylor, *Linear functionals on certain spaces of abstractly-valued functions*, *Annals of Mathematics*, (2), vol. 39 (1938), pp. 913-944.
6. N. Dunford, *Integration and linear operations*, these Transactions, vol. 40 (1936), pp. 474-494.
7. ———, *Uniformity in linear spaces*, these Transactions, vol. 44 (1938), pp. 305-356.
8. ———, *A mean ergodic theorem*, *Duke Mathematical Journal*, vol. 5 (1939), pp. 635-646.
9. N. Dunford and B. J. Pettis, *Linear operations among summable functions*, *Proceedings of the National Academy of Sciences*, vol. 25 (1939), pp. 544-550.
10. G. Fichtenholtz and L. Kantorovitch, *Sur les opérations linéaires dans l'espace des fonctions bornées*, *Studia Mathematica*, vol. 5 (1934), pp. 69-98.
11. M. Fréchet, *Les ensembles abstraits et le calcul fonctionnel*, *Rendiconti del Circolo Matematico di Palermo*, vol. 30 (1910), p. 19.
12. I. Gelfand, *Abstrakte Funktionen und lineare Operatoren*, *Recueil Mathématique*, vol. 4 (1938), pp. 235-284.
13. M. Gowerin, *Stieltjessche integration*, *Fundamenta Mathematicae*, vol. 27 (1936), pp. 254-268.
14. T. H. Hildebrandt, *On bounded linear functional operations*, these Transactions, vol. 36 (1934), pp. 868-875.
15. T. H. Hildebrandt and I. J. Schoenberg, *On linear functional operations and the moment problem for a finite interval in one or several dimensions*, *Annals of Mathematics*, (2), vol. 34 (1933), pp. 317-328.
16. Shizuo Kakutani, *Iteration of linear operations in complex Banach spaces*, *Proceedings of the Imperial Academy of Tokyo*, vol. 14 (1938), pp. 295-300.
17. L. Kantorovitch and B. Vulich, *Sur la représentation des opérations linéaires*, *Composito Mathematica*, vol. 5 (1937), pp. 119-165.
18. A. Kolmogoroff, *Über Kompaktheit der Funktionenmengen bei der Konvergenz im Mittel*, *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen*, 1931, pp. 60-63.
19. I. Maddaus, *On completely continuous linear transformations*, *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 279-282.
20. E. H. Moore and H. L. Smith, *A general theory of limits*, *American Journal of Mathematics*, vol. 44 (1922), pp. 102-121.
21. W. Orlicz, *Beiträge zur Theorie der Orthogonalentwicklungen II*, *Studia Mathematica*, vol. 1 (1929), pp. 241-255.
22. B. J. Pettis, *On integration in vector spaces*, these Transactions, vol. 44 (1938), pp. 277-304.

23. R. S. Phillips, *On integration in a linear convex topological space*, these Transactions, vol. 47 (1940), pp. 114-146.
24. J. Radon, *Theorie und Anwendung der absolut additiven Mengenfunktionen*, Sitzungsberichte der Akademie der Wissenschaften, Vienna, Class IIa, vol. 122 (1913), p. 1384.
25. M. Riesz, *Sur les ensembles compacts de fonctions sommables*, Acta Szeged, vol. 6 (1933), pp. 136-142.
26. J. Tamarkin, *On compactness of the space  $L^p$* , Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 79-84.
27. A. Tulajkov, *Zur Kompaktheit im Raum  $L^p$  für  $p=1$* , Nachrichten der Gesellschaft der Wissenschaften zu Göttingen, 1933, pp. 167-170.
28. Kôsaku Yosida, *Mean ergodic theorem in Banach spaces*, Proceedings of the Imperial Academy of Tokyo, vol. 14 (1938), pp. 292-294.

THE INSTITUTE FOR ADVANCED STUDY,  
PRINCETON, N. J.

## ON A TYPE OF ALGEBRAIC DIFFERENTIAL MANIFOLD

BY

J. F. RITT

The manifolds<sup>(1)</sup> to be investigated, which are manifolds of systems of differential polynomials in a single unknown, possess a degree of analogy to bounded sets of numbers. They are manifolds which may be said "not to contain infinity as a solution"; more definitely, zero is not a limit of reciprocals of solutions.

For manifolds of this type, which will be called *limited*, operations of addition, multiplication and differentiation will be studied. Given two manifolds<sup>(2)</sup>  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , their *arithmetic sum* is secured by completing into a manifold the totality of functions each of which is, in some area, the sum of a solution in  $\mathcal{M}_1$  and a solution in  $\mathcal{M}_2$ . Multiplication is defined similarly.

It turns out that if  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are general solutions of equations of the first order, and are limited, their sum and product are limited. On the other hand, as is shown by examples based on the theory of the elliptic functions, when  $\mathcal{M}_1$  and  $\mathcal{M}_2$  involve more than one arbitrary constant their limited character may not be communicated to their sum and product; what is equivalent to this, as far as multiplication is concerned, is the rather unexpected result that the product of two manifolds may contain zero even if neither manifold does.

The derivative of a limited manifold proves to be limited in all cases.

### LIMITED MANIFOLDS

1. Let  $\Sigma$  be a system of forms in the single unknown  $y$ . Let us suppose that  $\Sigma$  has solutions and that it has at least one solution which is not identically zero. The transformation  $z = 1/y$  carries every nonzero solution of  $\Sigma$  into a definite function  $z$ . There exist forms in  $z$  which vanish for every function  $z$  thus obtained. Let  $\Sigma'$  be the totality of such forms in  $z$ . It is not difficult to see that the manifold of  $\Sigma'$  is the set of the reciprocals of the nonzero solutions of  $\Sigma$ , enlarged perhaps by the adjunction of  $z=0$ .

If  $\Sigma'$  does not admit  $z=0$  as a solution, we shall call the manifold of the original system  $\Sigma$  *limited*<sup>(3)</sup>.

2. If  $\Sigma'$  has  $z=0$  as a solution,  $z=0$  cannot be an essential manifold for  $\Sigma'$ . If it were,  $\Sigma'$ , which is closed, would contain a form  $zA$  where  $A$  does not

Presented to the Society, September 12, 1940; received by the editors March 20, 1940.

(<sup>1</sup>) For indications in regard to the general theory to which this paper attaches, one may consult the author's paper in the second volume of the Semicentennial Publications of the American Mathematical Society.

(<sup>2</sup>) Not necessarily limited.

(<sup>3</sup>) If  $\Sigma$  admits only  $y=0$  as a solution, its manifold will also be called limited.

vanish for  $z=0$ . Now  $A$  would vanish for the reciprocal of every nonzero solution of the system  $\Sigma$ . It would thus be in  $\Sigma'$  and would rule out the solution  $z=0$ .

Thus, if the manifold of  $\Sigma$  is not limited, there is a dense set of values of  $x$  such that, given any point  $a$  of the set, any positive integer  $m$  and any  $\epsilon > 0$ , we can find a solution of  $\Sigma$  whose reciprocal is analytic at  $a$  and has a Taylor expansion at  $a$  in which the first  $m+1$  coefficients have moduli less than  $\epsilon$ . When the manifold of  $\Sigma$  is limited, no point exists which has the property, just stated, of the points  $a$ .

3. Let  $\Sigma$  be a closed system of forms in  $y$  which admits solutions. We shall prove that *for the manifold of  $\Sigma$  to be limited, it is necessary and sufficient that  $\Sigma$  contain a form  $A$  which, considered as a polynomial in  $y$  and its derivatives, possesses a term in  $y$  alone, that is, a term free of the  $y_i$  with  $i > 0$ , which is of higher degree than every other term in  $A$ .*

Let the manifold be limited. We may suppose that there are solutions other than  $y=0$ . Then  $\Sigma'$ , as above, contains a form  $1+K$  with  $K$  a nonzero form which vanishes for  $z=0$ . Making the substitution  $z=1/y$  in  $K$ , and clearing fractions, we obtain a form in  $\Sigma$  answering to the description of  $A$ .

Conversely, let  $\Sigma$  contain a form  $A$  as described. If we put  $y=1/z$  in  $A$  and clear fractions, we secure a form  $B$  in  $\Sigma'$ , one of whose terms, free of proper derivatives of  $z$ , is of lower degree than every other term. Thus, if  $z=0$  were in the manifold of  $\Sigma'$ , it would be an essential manifold. This, by §2, is impossible.

#### CONSIDERATIONS OF GENERAL THEORY

4. We present here a theorem of a general character which will be employed in §10.

Let  $\Sigma$  be a nontrivial closed irreducible system in the unknowns  $u_1, \dots, u_q; y_1, \dots, y_p$  with the  $u_i$  (which may be nonexistent) arbitrary and with  $p > 1$ . Let  $m$  be any positive integer not greater than  $p$ . Those forms in  $\Sigma$  which involve only the  $u_i$  and  $y_1, \dots, y_m$  constitute a closed irreducible system  $\Sigma_m$  in the unknowns just mentioned. For  $m=p$ ,  $\Sigma_m$  is  $\Sigma$ .

Let  $m < p$ . Given a solution

$$(1) \quad u_i; y_1, \dots, y_m$$

of  $\Sigma_m$ , analytic in an area  $\mathfrak{A}_1$ , there may exist an area  $\mathfrak{A}_2$  contained in  $\mathfrak{A}_1$  and a set of functions

$$(2) \quad y_{m+1}, \dots, y_p,$$

analytic in  $\mathfrak{A}_2$ , such that (1) and (2) constitute a solution of  $\Sigma$  in  $\mathfrak{A}_2$ . In that case, we shall say that the solution (1) of  $\Sigma_m$  can be *completed* into a solution of  $\Sigma$ .

We are going to prove that *there exists a form  $G$  in  $u_1, \dots, u_q; y_1, \dots, y_m$*

which does not belong to  $\Sigma_m$  and which has the property that every solution of  $\Sigma_m$  which does not annul  $G$  can be completed into a solution of  $\Sigma$ .

5. Let

$$(3) \quad A_1, \dots, A_p$$

be a basic set of  $\Sigma$ ,  $A_i$  introducing  $y_i$ . Let the order of  $A_i$  in  $y_i$  be  $r_i$ . Let  $S_i$  and  $I_i$  be respectively the separant and initial of  $A_i$ .

We consider the system of forms

$$(4) \quad A_1, \dots, A_{m+1},$$

which is a basic set of  $\Sigma_{m+1}$ . Let a form  $L$  be given which is not in  $\Sigma_{m+1}$  and which is such that every  $y_{ij}$  appearing in  $L$  has  $i \leq m+1$  and  $j \leq r_i$ . We place no restrictions on the  $u_{ij}$  in  $L$ . We shall establish a relation

$$(5) \quad R = M + NLS_{m+1}$$

of the following description. The  $y_{ij}$  in  $R$ ,  $M$  and  $N$  have  $i \leq m+1$  and  $j \leq r_i$ .  $M$  is contained in  $\Sigma_{m+1}$ .  $R$ , distinct from zero, is free of the  $y_{ir_i}$ . Thus  $R$  is not in  $\Sigma_{m+1}$ .

6. Let (4) be considered as a set of simple forms. Then (4) will be a basic set of a prime system<sup>(4)</sup>  $\Pi$ . Now  $LS_{m+1}$  (simple form) is not in  $\Pi$ . Then every indecomposable system held by  $\Pi + LS_{m+1}$  has fewer unconditioned unknowns than  $\Pi$ . There exists thus a relation (5) with all forms simple forms,  $R$  being distinct from zero and free of the  $y_{ir_i}$ , and  $M$  belonging to  $\Pi$ . It remains only to consider the forms in (5) as differential polynomials.

7. Let

$$J = S_1 \dots S_m I_1 \dots I_{m+1} R.$$

Let  $J$  be considered as a polynomial in the  $y_{m+1,j}$ , with coefficients which are forms in the  $u_i$  and  $y_1, \dots, y_m$ . Not all of these coefficients can be in  $\Sigma_m$ . If they were,  $J$  would be in  $\Sigma_{m+1}$ . Let  $H$  be a coefficient which is not in  $\Sigma_m$ .

We say that any solution  $\hat{u}_i; \hat{y}_1, \dots, \hat{y}_m$  of  $\Sigma_m$  which does not annul  $H$  can be completed into a solution of  $\Sigma_{m+1}$  which does not annul  $L$ .

Let  $a$  be a value of  $x$  at which all functions of  $x$  which we shall use are analytic and at which the above solution of  $\Sigma_m$  does not annul  $H$ . Let the solution be substituted into  $J$  and let numerical values then be attributed to the  $y_{m+1,j}$  with  $j < r_{m+1}$  in such a way as to give  $J$  a numerical value, for  $x=a$ , which is not zero. We can then find a numerical value for the  $r_{m+1}$ th derivative of  $y_{m+1}$  which, together with  $x=a$ , etc., annuls  $A_{m+1}$ . Referring to  $M$  in (5), we see that, because the remainder of  $M$  with respect to (4) is zero and  $I_1, \dots, I_{m+1}$  do not vanish for the indicated numerical values, the values cause  $M$  to vanish. Hence  $LS_{m+1}$  does not vanish for the values. This means

<sup>(4)</sup> The  $u_{ij}$  in  $\Pi$  are those appearing in (4) and in  $L$ .



that the above solution of  $\Sigma_m$  can be completed into a regular solution of (4) which does not annul  $L$ , so that our statement is proved. We note that the  $y_{ij}$  in  $H$  have  $j \leq r_i$ ,  $i = 1, \dots, m$ .

8. We might have taken  $L = 1$  in §7. On this basis, let  $K$ , a form not in  $\Sigma_{m+1}$  which involves only such  $y_{ij}$  as appear in (4), be such that every solution of  $\Sigma_{m+1}$  which does not annul  $K$  can be completed into a solution of  $\Sigma_{m+2}$ . If, returning to  $\Sigma_m$ , we take  $L = K$ , we find an  $H$  such that the solutions of  $\Sigma_m$  which do not annul  $H$  can be completed into solutions of  $\Sigma_{m+2}$ . The proof of the theorem stated in §4 is thus easy to conclude.

#### SUMS, PRODUCTS AND DERIVATIVES

9. Let  $\Sigma_1$  and  $\Sigma_2$  be systems of forms in  $y$ , each system possessing solutions. It is possible to form, in various ways, sums  $y' + y''$  where  $y'$  and  $y''$ , solutions respectively of  $\Sigma_1$  and of  $\Sigma_2$ , have the same area of analyticity. The manifold of the system of those forms in  $y$  which vanish for all sums  $y' + y''$  will be called the *arithmetic sum* of the manifolds of  $\Sigma_1$  and  $\Sigma_2$ . We define similarly *arithmetic product*, using all products  $y'y''$ .

Let  $\Sigma$  be a system of forms in  $y$  which possesses solutions. There are forms in  $y$  which vanish if  $y$  is the derivative of any solution of  $\Sigma$ . The manifold of the totality of such forms will be called the *derivative* of the manifold of  $\Sigma$ .

*Examples.* If  $\Sigma_1$  and  $\Sigma_2$  are the forms  $y_1 - 1$  and  $xy_1 - y$  respectively, the arithmetic sum of their manifolds is the two-parameter family of functions  $y = ax + b$ . The arithmetic product is the family  $ax^2 + bx$ . The derivative of the manifold of  $\Sigma_2$  is the manifold of  $y_1$ .

10. Certain solutions in the sum of two manifolds may not be sums  $y' + y''$ . Such special solutions will now be examined.

Let  $\Sigma_1$  be a nontrivial closed system of forms in the unknown  $u$ . Let  $\Sigma_2$  be a similar system in  $v$ . Let  $\Lambda$  be a system in  $u, v, y$  consisting of the forms in  $\Sigma_1$ , those in  $\Sigma_2$  and  $y - (u + v)$ . Let  $\Omega$  be the totality of forms in  $u, v, y$  which hold  $\Lambda$ . One can prove that  $\Omega$  contains nonzero forms in  $y$  alone. Let  $\Sigma'$  be the totality of forms in  $\Omega$  which are free of  $u$  and  $v$ . If  $\Omega$  is the intersection of closed irreducible systems  $\Omega_1, \dots, \Omega_r$ , then  $\Sigma'$  will be the intersection of those subsystems of the  $\Omega_i$  which are free of  $u$  and  $v$ .

We refer now to §4. We see that there is a nonzero form  $G$  in  $y$  alone, holding no essential irreducible manifold in the manifold of  $\Sigma'$ , which is such that every solution of  $\Sigma'$  which does not annul  $G$  can be represented, in some area, as the sum of a solution of  $\Sigma_1$  and a solution of  $\Sigma_2$ .

Let us apply these conclusions to the systems  $\Sigma_1$  and  $\Sigma_2$  of §9, which we shall suppose closed and nontrivial, with the respective manifolds  $\mathcal{M}_1$  and  $\mathcal{M}_2$  of sum  $\mathcal{M}$ . Let  $G$  be a form in  $y$ , holding no essential irreducible manifold in  $\mathcal{M}$ , which is such that every solution in  $\mathcal{M}$  which does not annul  $G$  is the sum of solutions taken from  $\mathcal{M}_1$  and  $\mathcal{M}_2$ .

Let  $\bar{y}$  be any solution in  $\mathfrak{M}$ . Let  $\mathfrak{A}'$  be any area in which  $\bar{y}$  is analytic. Let  $m$  be a positive integer and  $\epsilon$  a positive number. Let  $\bar{y}$  be a solution in  $\mathfrak{M}$ , analytic in an area  $\mathfrak{A}_1$  contained in  $\mathfrak{A}'$ ,  $\bar{y}$  being so taken that  $G$  is not annulled by  $\bar{y}$  at any point of  $\mathfrak{A}_1$  and that  $\bar{y} - \bar{y}$  has at each point of  $\mathfrak{A}_1$  a Taylor expansion in which the first  $m+1$  coefficients are of moduli less than  $\epsilon$ . The existence of  $\bar{y}$  is obvious. Let  $\mathfrak{A}'_1$  be an area contained in  $\mathfrak{A}_1$  in which  $\bar{y}$  is the sum of solutions taken from  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ . We now find a second  $\bar{y}$ , using an area  $\mathfrak{A}_2$  in  $\mathfrak{A}'_1$ , a larger  $m$  and a smaller  $\epsilon$ . Continuing, we see that *there exists a set of points, dense in the area in which  $\bar{y}$  is analytic, such that, given any point  $a$  of the set, any positive integer  $m$  and any  $\epsilon > 0$ , there is a solution  $\bar{y}$  in  $\mathfrak{M}$  which, for the neighborhood of  $a$ , is the sum of solutions taken from  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ , the first  $m+1$  coefficients in the expansion of  $\bar{y} - \bar{y}$  at  $a$  being of moduli less than  $\epsilon$ .*

A similar result holds for the product of  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ .

#### DESCRIPTION OF RESULTS OF PAINLEVÉ

11. In §12 we shall employ results of Painlevé concerning the algebroid character of the solution of an algebraic differential equation of the first order<sup>(6)</sup>. While these results have received enough attention to warrant describing them as classic, they have not thus far, to our knowledge, been given didactic exposition. Here, we shall limit ourselves to formulating Painlevé's results in a manner which will permit us to employ them with precision<sup>(6)</sup>.

Let  $A$  be an algebraically irreducible form in  $y$  of the first order, of degree  $n$  in  $y_1$ . Let  $\mathfrak{A}$  be the area in which the coefficients in  $A$  are meromorphic. There figures, in the statement of Painlevé's results, a set of points  $\mathcal{E}$ , contained in  $\mathfrak{A}$ , which includes the poles of the coefficients in  $A$  and has no limit point in the interior of  $\mathfrak{A}$ . When  $\mathcal{E}$  is removed from  $\mathfrak{A}$ , there remains an open region  $\mathfrak{A}'$ .

Let  $x_0$  be any point of  $\mathfrak{A}'$ . Let  $b$  be any finite number. Then, given any number  $y_0$ , close to  $b$  and distinct from  $b$ ,  $A$  has precisely  $n$  distinct solutions analytic at  $x_0$  and assuming the value  $y_0$  at  $x_0$ .

There exist, furthermore, a certain number  $j$  (depending on  $x_0$  and  $b$ ) of equations

$$(6) \quad y^{m_i} + \alpha_{1i}(x, y_0)y^{m_i-1} + \cdots + \alpha_{m_i i}(x, y_0) = 0, \quad i = 1, \dots, j,$$

whose descriptions and roles are as follows. The  $\alpha_{ki}$  are functions of  $x$  and  $y_0$ , analytic for  $|x - x_0| < \delta$ ,  $|y_0 - b| < \delta$ , where  $\delta$  is some positive number depending on  $x_0$  and  $b$ . For  $y_0$  close to  $b$  and distinct from  $b$ , each of the  $n$  solu-

<sup>(6)</sup> Painlevé, *Leçons sur la Théorie des Équations Différentielles Professées à Stockholm*, Paris, 1897, pp. 70-76.

<sup>(6)</sup> The matter is not very difficult to work out, starting with the indications given by Painlevé. It is helpful to read Schlesinger, *Gewöhnliche Differenzialgleichungen*, Chapter 3, where somewhat related questions are considered. The Weierstrass preparation theorem can be employed to advantage.

tions of  $A$  mentioned above satisfies one of the equations (6) in the neighborhood of  $x = x_0$ . Furthermore, every solution in the general solution<sup>(7)</sup> of  $A$  which is analytic at  $x_0$  and assumes the value  $b$  at  $x_0$  satisfies one of the equations (6). Again, if  $y(x)$  is a function analytic in an area contained in  $|x - x_0| < \delta$  and if  $y(x)$  satisfies one of the equations (6) with  $y_0$  fixed at a value interior to a circle of center  $b$  and radius  $\delta$ , then  $y(x)$  is a solution in the general solution of  $A$ . For a given  $y_0$  close to  $b$ , (6) may yield, in addition to solutions of  $A$  which equal  $y_0$  at  $x_0$ , other solutions of  $A$  analytic at  $x_0$ .

We now deal with solutions of  $A$  which assume large values at  $x_0$ . There exists a  $g > 0$  such that, for  $|y_0| > g$ ,  $A$  has precisely  $n$  distinct solutions, analytic at  $x_0$ , which assume the value  $y_0$  at  $x_0$ . There exists a number  $h$  (independent of  $x_0$ ) of equations

$$(7) \quad z^{pi} + \beta_{1i}(x, z_0)z^{pi-1} + \cdots + \beta_{pi}(x, z_0) = 0, \quad i = 1, \dots, h,$$

with  $\beta_{ki}$  which are analytic for  $x = x_0, z_0 = 0$ . Given any solution  $y$  of  $A$ , analytic at  $x_0$  and assuming there a large value  $y_0$ , the function  $z = 1/y$  satisfies one of the equations (7) with  $z_0 = 1/y_0$ . Given any function  $z$  distinct from zero, obtained from the equations (7) for a small value of  $z_0$ , the reciprocal of  $z$  is a solution of  $A$ .

We proceed to apply these results of Painlevé.

#### LIMITED SUMS AND PRODUCTS

12. We prove the following theorem.

**THEOREM I.** *Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be general solutions of forms of the first order in  $y$ . Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be limited. Then the sum and the product of  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are limited manifolds.*

We take first the case of the product, disposing of that case by establishing the following result<sup>(8)</sup>.

**THEOREM II.** *Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be general solutions of forms of the first order. Let neither  $\mathcal{M}_1$  nor  $\mathcal{M}_2$  have zero among its solutions. Then zero is not a solution in the product of  $\mathcal{M}_1$  and  $\mathcal{M}_2$ .*

Let us assume that zero is in the product. There are values of  $x$  at which zero can be approximated, as in §10, by products of solutions in  $\mathcal{M}_1$  and  $\mathcal{M}_2$ . We select a value  $x_0$  of this type which does not belong to either of the sets  $\mathcal{E}$

(7) The notion of general solution, as employed here, does not, of course, appear in Painlevé's work.

(8) What is involved here is the following. Let  $\mathcal{M}$  be the product of the limited  $\mathcal{M}_1$  and  $\mathcal{M}_2$  of Theorem I. By §1, the reciprocals of the solutions distinct from zero in  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are manifolds. We represent the manifolds of reciprocals, which are seen without trouble to be general solutions of forms of the first order, by  $\mathcal{M}'_1$  and  $\mathcal{M}'_2$  and their product by  $\mathcal{M}'$ . A form  $F$  holds  $\mathcal{M}'$  if it vanishes for every  $1/(y'y'')$  with  $y'$  in  $\mathcal{M}_1$  and  $y''$  in  $\mathcal{M}_2$ . By §10,  $F$  will vanish for the reciprocal of every nonzero solution in  $\mathcal{M}$ . Thus, if  $\mathcal{M}$  is not limited,  $\mathcal{M}'$  contains zero.

of §11 associated with the forms whose general solutions are  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ .

For convenience, we use  $y$  to designate solutions in  $\mathfrak{M}_1$  and  $u$ , similarly, for  $\mathfrak{M}_2$ . Let there be given a sequence of  $yu$  whose expansions tend toward zero at  $x_0$ . From this sequence we can select a subsequence in which the values  $y(x_0)$ ,  $u(x_0)$  tend toward definite limits, finite or infinite. We may thus, and shall, assume that such limits exist for the given sequence. We assume, as we may, that the limit of the  $y(x_0)$  is zero. The limit of the  $u(x_0)$  will be a quantity  $c$ , finite or infinite.

We treat first the case in which  $c$  is finite.

We may suppose that all of the  $y$  satisfy a single equation (6). We write this equation here in the form

$$(8) \quad y^m + \alpha_1(x, y_0)y^{m-1} + \cdots + \alpha_m(x, y_0) = 0.$$

Similarly, the  $u$  may be supposed to satisfy an equation

$$(9) \quad u^n + \beta_1(x, u_0)u^{n-1} + \cdots + \beta_n(x, u_0) = 0.$$

Because zero is not a solution in  $\mathfrak{M}_1$ ,  $\alpha_m$  cannot vanish identically in  $x$  for a small value of  $y_0$ ; similarly,  $\beta_n$  cannot vanish in  $x$  for a value of  $u_0$  close to  $c$ .

The theory of symmetric functions shows that the  $yu$  satisfy an equation

$$(10) \quad (yu)^{mn} + \gamma_1(yu)^{mn-1} + \cdots + \gamma_{mn} = 0$$

where the  $\gamma$  are polynomials in the  $\alpha$  and the  $\beta$ , with  $\gamma_{mn} = \alpha_m^n \beta_n^m$ . Because the Taylor expansions of the  $yu$  approach zero, (10) must be satisfied, for  $y_0 = 0$ ,  $u_0 = c$ , by  $yu = 0$ . This is not so. We have thus disposed of the case finite.

Now, suppose that  $c = \infty$ . We let  $z$  represent the reciprocals of the  $u$ . We may assume that the  $z$  all satisfy an equation

$$(11) \quad z^p + \cdots + \delta_p(x, z_0) = 0.$$

Then the  $y/z$  satisfy an equation

$$(12) \quad \phi_0(y/z)^{mp} + \cdots + \phi_{mp} = 0$$

with  $\phi$  which are polynomials in the  $\alpha$  and  $\delta$ , and with, in particular,

$$\phi_0 = \delta_p^m, \quad \phi_{mp} = \alpha_m^p.$$

We reach the contradiction that (12) is satisfied by  $y/z = 0$  for  $y_0 = z_0 = 0$ . This concludes the proof of our statement in regard to products.

Continuing with Theorem I, we consider the limited  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$ , under the assumption that their sum is not limited.

Using  $y$  for  $\mathfrak{M}_1$  and  $u$  for  $\mathfrak{M}_2$ , we consider an  $x_0$ , and a sequence of  $y+u$  for which the expansions of the  $1/(y+u)$  tend toward the expansion of zero at  $x_0$ (<sup>9</sup>). We shall assume, furthermore, that the sequences of values  $y(x_0)$ ,

(<sup>9</sup>) That such a sequence exists can be shown without difficulty by the method of §10.

$u(x_0)$  tend toward definite limits, finite or infinite. At least one of these limits is infinite. Let this be so for the  $u(x_0)$ . We suppose first that the  $y(x_0)$  have a finite limit.

We arrange so that the  $y$  satisfy an equation (8) and the reciprocals  $z$  of the  $u$  an equation (11). Let

$$w = \frac{1}{y+u} = \frac{z}{1+yz}.$$

We find the  $w$  to satisfy an equation  $\phi_0 w^{m_p} + \dots + \phi_{m_p} = 0$  with  $\phi_{m_p} = \delta_p^m$ . We must thus have  $\delta_p(x, 0) = 0$ . This produces the contradiction that  $\mathfrak{M}_2$  is not limited. The case in which the  $y(x_0)$  approach  $\infty$  is handled in much the same way.

#### EQUATIONS OF HIGHER ORDER

13. We shall show by means of examples suggested by the theory of the elliptic functions that the above results cannot be extended to equations of the second order.

The nonconstant solutions of

$$(13) \quad y_1^2 = 4(y^3 - e^3),$$

where  $e$  is any constant, satisfy the equation

$$y_2 - 6y^2 = 0,$$

whose manifold  $\mathfrak{M}_1$  is, by §3, limited. If, in (13), we replace  $y$  by  $y+e$ , (13) goes over into

$$(14) \quad y_1^2 = 4(y^3 + 3ey^2 + 3e^2y)$$

which implies, when  $y$  is not a constant,

$$(15) \quad 16y^6 - 8y^3y_1^2 - 8y^4y_2 + y_1^4 - 4yy_1^2y_2 + 4y^2y_2^2 = 0$$

with a limited manifold  $\mathfrak{M}_2$ . For any constant  $e$ , arbitrarily large, there are solutions in  $\mathfrak{M}_1$  and  $\mathfrak{M}_2$  respectively whose difference is  $e$ . This is enough to show that the theorem on sums does not hold for equations of the second order.

If in (14), we replace  $y$  by  $3e^2/y$ , (14) remains invariant. Thus, for any  $e$ , (15) has two solutions whose product is  $3e^2$ . The theorem on the product thus does not carry over to the second order.

#### THE DERIVATIVE

14. We prove the following theorem.

**THEOREM.** *The derivative of a limited manifold is limited.*



Let  $\mathfrak{M}$ , limited, be held by a form  $F = y^p - G$  with every term in  $G$  of degree less than  $p$ . We have  $y^p \equiv G, (F)$ . Now  $y_1^{2p-1} \equiv 0, (y^p)$ . Hence there is a relation  $y_1^{2p-1} \equiv H, (F)$  with every term in  $H$  of degree less than  $2p-1$ .

We arrange  $y^p - G$  and  $y_1^{2p-1} - H$  in powers of  $y$ , securing two polynomials in  $y$ ,

$$(16) \quad A_0 y^p + \cdots + A_p$$

and

$$(17) \quad B_0 y^q + \cdots + B_q.$$

Here  $A_0 = 1$  and  $A_i$  is of degree less than  $i$  for  $i > 0$ . Also,  $B_q$  has  $y_1^{2p-1}$  as one term and its other terms are of degree less than  $2p-1$ . Each  $B_i$  with  $i < q$  is of degree less than  $2p-q+i-1$ .

We consider the resultant,  $R$ , of (16) and (17) with respect to  $y$ . One of the terms of  $R$  is  $A_0^q B_q^p$ , that is,  $B_q^p$ . Now  $B_q^p$  contains  $y_1^{(2p-1)p}$  and its other terms are of degree less than  $(2p-1)p$ . Consider any other term in  $R$ ,

$$T = k A_{\mu_1} \cdots A_{\mu_q} B_{\nu_1} \cdots B_{\nu_p}.$$

We have

$$\mu_1 + \cdots + \mu_q + \nu_1 + \cdots + \nu_p = pq.$$

At least one  $\mu$  is positive and at least one  $\nu$  is less than  $p$ . We have thus, for the degree  $d$  of  $T$  in the  $y_i$ ,

$$d < \sum \mu_i + \sum (2p - q + \nu_i - 1) = (2p - 1)p.$$

Thus  $R = y_1^{(2p-1)p} + K$  where each term in  $K$  is of degree less than  $(2p-1)p$ . We note that the  $y_i$  in  $K$  have  $i > 0$ . Now  $R$  holds  $\mathfrak{M}$ . Then the derivative of  $\mathfrak{M}$  is held by the form obtained from  $R$  by replacing each  $y_i$  appearing in  $R$  by  $y_{i-1}$ . Thus the derivative of  $\mathfrak{M}$  is limited.

15. We shall prove that, if  $F$ , in §14, is of the first order, the derivative of  $\mathfrak{M}$  is held by a form  $y^q + L$  with  $L$  of the *first* order and of degree less than  $q$ .

It will suffice to prove that the derivative of the manifold of  $F$  is held by a form  $y^q + L$  as just described. In that proof, we may and shall assume that  $F$  is algebraically irreducible.

We consider  $F$  and its derivative  $F_1$  as polynomials in  $y$  and denote their resultant with respect to  $y$ , which is not identically zero, by  $R$ . Now  $R$  must involve  $y_2$  effectively; otherwise  $R$ , which holds  $F$ , would be divisible by  $F$ , which involves  $y$ .

We show now that  $R$  contains a term in  $y_1$  alone which is of higher degree than any other term in  $R$ . This will prove our statement.

Let us assume that the terms of highest degree in  $R$  involve  $y_2$ . We consider the equation  $R=0$  as an algebraic equation for  $y_2$ . It will be satisfied,



for the neighborhood of  $y_1 = \infty$ , by some series of descending rational powers of  $y_1$ ,

$$(18) \quad y_2 = \alpha y_1^{s/r} + \beta y_1^{(s-1)/r} + \dots$$

with  $r > 0$ ,  $s \leq r$  and  $\alpha, \beta$ , etc., functions of  $x$  analytic in some area<sup>(16)</sup>.

We substitute this expression for  $y_2$  into  $F_1$ , whereupon  $F_1$  goes over into a polynomial  $f$  in  $y$  whose coefficients are infinite series in  $y_1$ . We consider the equations  $F=0$  and  $f=0$  as algebraic equations for  $y$ . They must have a common solution given by a series of descending powers of  $y_1$ , effectively involving  $y_1$ ,

$$(19) \quad y = \phi y_1^{t/u} + \dots$$

with  $u$  a multiple of  $r$ , (19) converging for large values of  $y_1$ . We assume that  $\phi \neq 0$ .

Suppose first that  $t > 0$ . Then  $t < u$  since, when  $F$  is considered as a polynomial in  $y$  and  $y_1$ , its term  $y^p$  is of higher degree than every other term. From (19), we find by inversion, for the neighborhood of  $y = \infty$ , a series of descending powers for  $y_1$  of the type

$$(20) \quad y_1 = \psi y^{u/t} + \dots$$

Substituting  $y_1$ , as in (20), into (18), we find a series for  $y_2$

$$(21) \quad y_2 = \lambda y^{s u / r t} + \dots$$

If we replace  $y_1$  and  $y_2$  in  $F_1$  by their expressions in (20) and (21),  $F_1$  will vanish identically in  $x$  and  $y$ . But if we replace  $y_1$  in the equation  $F_1=0$  by the second member of (20) and solve the resulting equation for  $y_2$ , we will find for  $y_2$  a series in  $y$  obtained by differentiating the second member of (20) and replacing  $y_1$  in the result by its expression in (20). The series thus obtained begins effectively with a power of  $y$  whose exponent is  $(2u/t) - 1$ , which exceeds the first exponent in (21).

Now suppose that  $t=0$ . Then (19) yields an expansion for  $y_1$  in *ascending* powers of  $y - \phi$  of the type

$$(22) \quad y_1 = \mu(y - \phi)^{-k} + \dots$$

where  $k$  is a positive rational number. Substituting (22) into (18) and proceeding as above, we find again a contradiction of the fact that  $s < r$ .

The case of  $t < 0$  is handled in the same way.

16. If  $F$  is of order  $r > 1$ , we cannot infer that the derivative of  $\mathfrak{M}$  is held by a form  $y^s + L$  with  $L$  of order not higher than  $r$  and of degree less than  $q$ . Let  $\mathfrak{M}$  be the manifold of  $y_2 - y^2$ . We find that  $\mathfrak{M}$  is held by

<sup>(16)</sup> The second member of (18) may be zero.

$$A = y_2^2 - 4y_2y_1.$$

Suppose now that  $\mathfrak{M}$  is held by  $B = y_1^q + K$  where  $K$ , free of  $y$ , is of order not higher than 3 and has no term of degree as high as  $q$ . Because  $y_1^q$  is effectively a term in  $B$  and is divisible neither by  $y_2$  nor  $y_1$ ,  $B$  is not divisible by  $A$ . Hence, the resultant  $R$  of  $A$  and  $B$  with respect to  $y_2$  is a nonzero polynomial  $R$  in  $y_1$  and  $y_2$ . Putting  $y_2 = y^2$  in  $R$ , we find the contradiction that  $\mathfrak{M}$  is held by a form of order less than 2.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

